

# Introducción a los Modelos Lineales: fundamentos teóricos y prácticos

Javier Fernández-López, Profesor Ayudante Doctor  
Unidad de Matemática Aplicada  
Departamento de Biodiversidad, Ecología y Evolución, UCM

# Sobre mi...



hunting n-mixture mammal  
data-action glm nimble poisson  
regression rstat abundance enetwild  
iberia distribution binomial fungi occupancy  
biogeography monitoring wildboar rabbit migration  
connectivity models bayesian  
biodiversity fox survey cartography netlogo  
intheria deer maps carnivores plants  
hierarchical-models abm iberconejo  
birds camera-trap

Javier Fernández-López  
javfer05@ucm.es



REAL JARDÍN  
BOTÁNICO



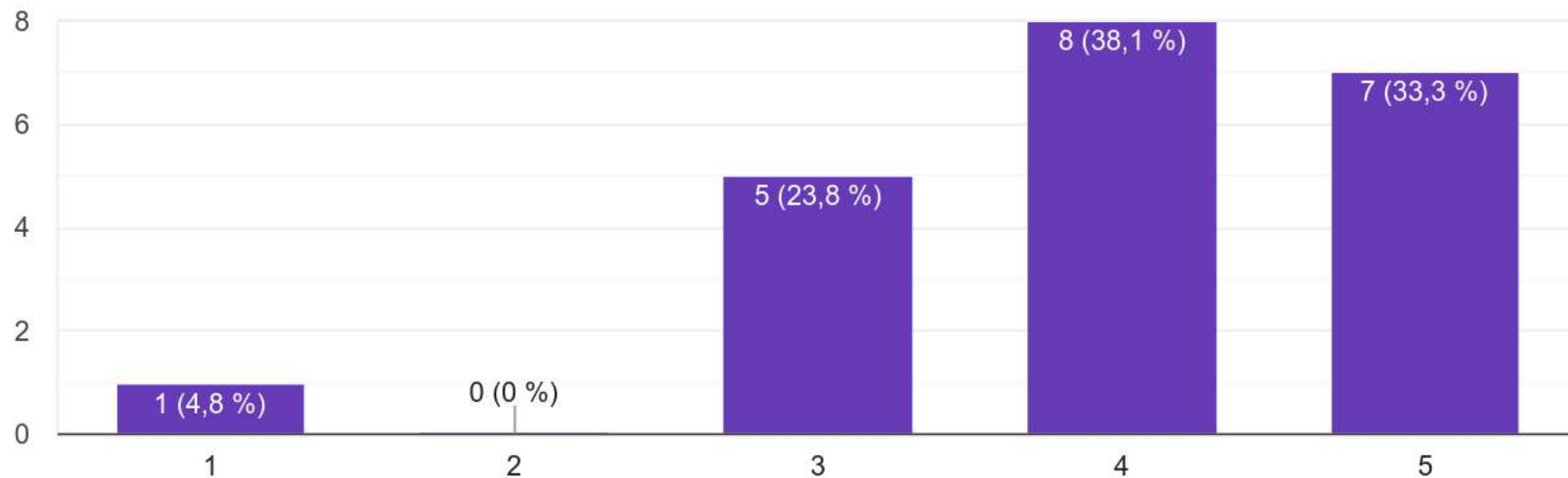
CENTRE D'ÉCOLOGIE  
FONCTIONNELLE  
& EVOLUTIVE

# Sobre vosotrxs...

¿Cómo de familiarizad@ estás con el lenguaje R?

 Copiar gráfico

21 respuestas

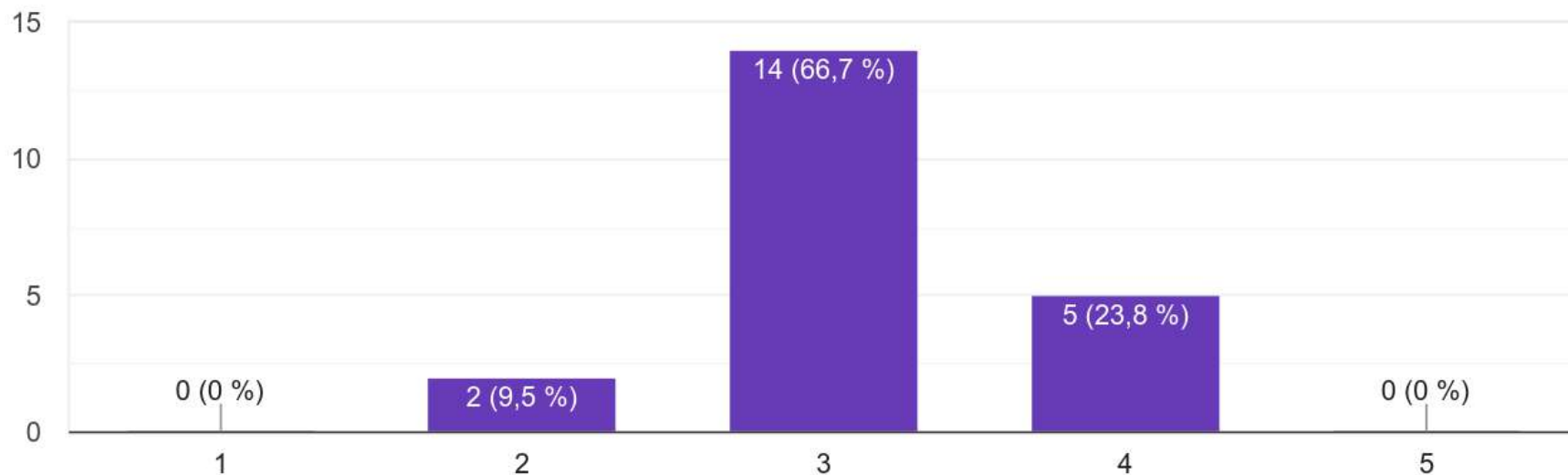


# Sobre vosotrxs...

¿Cómo de familiarizad@ estás con los modelos lineales (regresiones, GLM, etc.)?

 [Copiar gráfico](#)

21 respuestas



## Sobre vosotrxs...

Grupo muy diverso (tres programas de doctorado):

50% BFGE

20% CSA

23.3% ABA

6.7% otros.

Muchas disciplinas diferentes: ganadería, agricultura, ecología, microbiología, calidad alimentaria, nutrición, botánica, zoología...

Reto, pero también oportunidad!

## Sobre el curso...

- Módulo 1 - Introducción a la modelización  
(5 horas, 20 y 27 de octubre)

# Sobre el curso...

- Módulo 1 - Introducción a la modelización  
(5 horas, 20 y 27 de octubre)
- Módulo 2 - Modelos lineales generalizados  
(5 horas, 3 y 10? de noviembre)

# Sobre el curso...

- Módulo 1 - Introducción a la modelización  
(5 horas, 20 y 27 de octubre)
- Módulo 2 - Modelos lineales generalizados  
(5 horas, 3 y 10? de noviembre)
- Módulo 3 - Modelos mixtos y otras extensiones  
(5 horas, 17? y 24 de noviembre)



## Sobre el curso...

- Dinámica. Teoría y práctica (y Paint)

# Sobre el curso...

- Dinámica. Teoría y práctica (y Paint)
- Presentaciones, material y grabación:

<https://jabiologo.github.io/web/tutorials/estadisticaUPV.html>

# Sobre el curso...

- Dinámica. Teoría y práctica (y Paint)
- Presentaciones, material y grabación:

<https://jabiologo.github.io/web/tutorials/estadisticaUPV.html>

- Dudas y preguntas: estadística (interrumpir?) y software (esperar?)

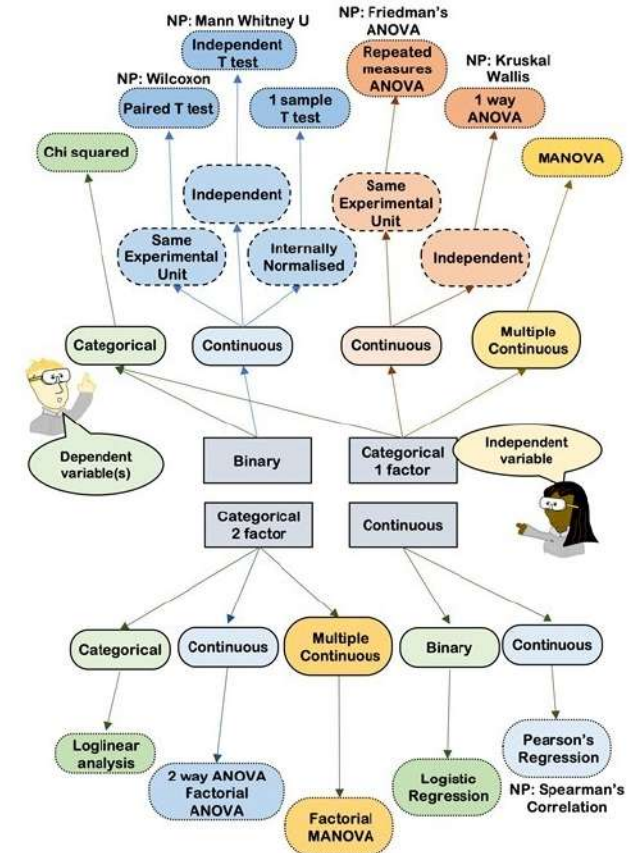
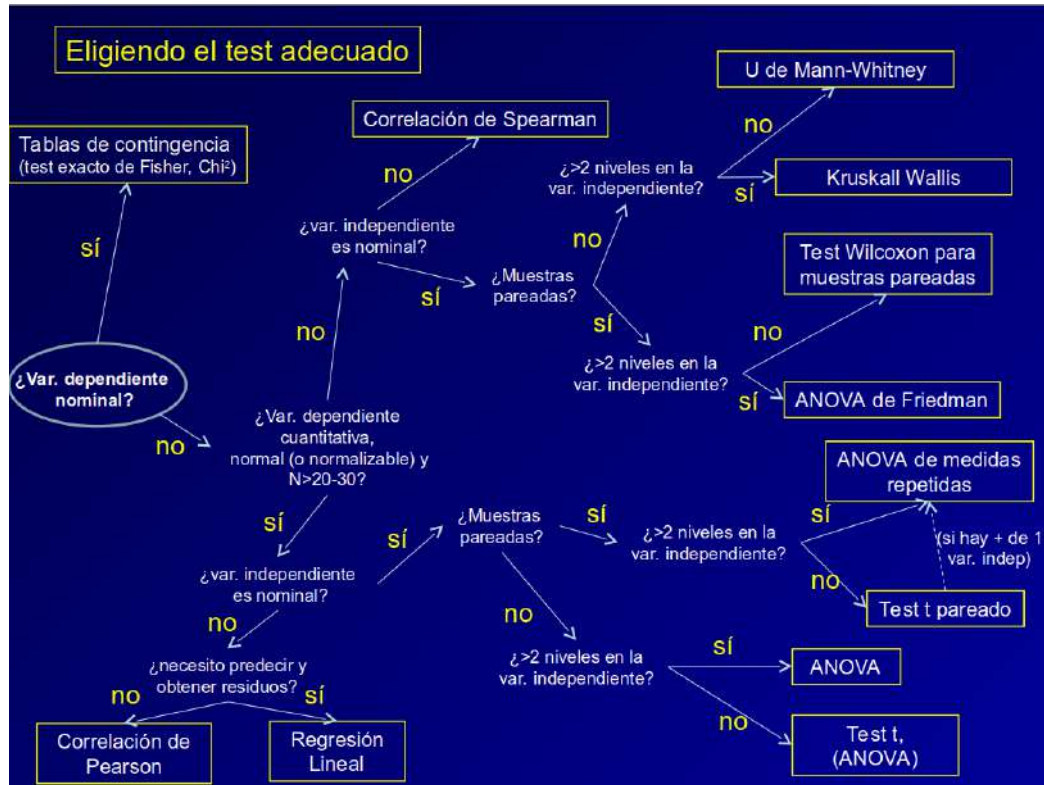
# Sobre el curso...

- Dinámica. Teoría y práctica (y Paint)
- Presentaciones, material y grabación:

<https://jabilologo.github.io/web/tutorials/estadisticaUPV.html>

- Dudas y preguntas: estadística (interrumpir?) y software (esperar?)
- Filosofía docente

# Sobre el curso...



# Bloque 1: Introducción a la modelización

- El método científico y la estadística
- Diseños experimentales VS datos observacionales
- Concepto de modelo
- Las variables y las distribuciones de probabilidad
- Las simulaciones como un laboratorio
- Introducción al modelo lineal

# Bloque 1: Introducción a la modelización

- El método científico y la estadística
- Diseños experimentales VS datos observacionales
- Concepto de modelo
- Las variables y las distribuciones de probabilidad
- Las simulaciones como un laboratorio
- Introducción al modelo lineal

# El método científico

Observación → Hipótesis



Experimentación

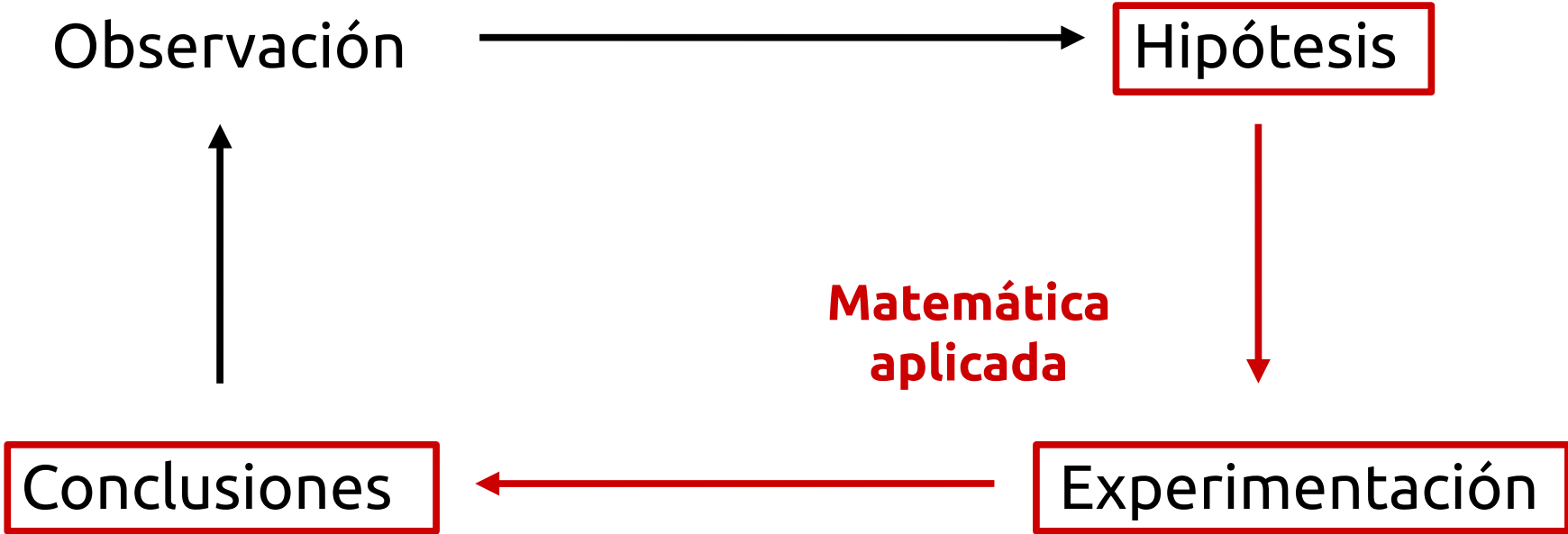


Conclusiones

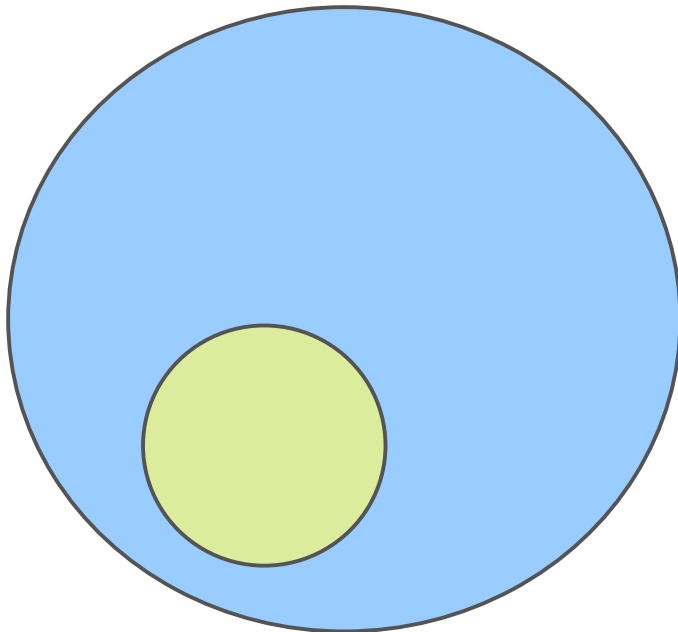




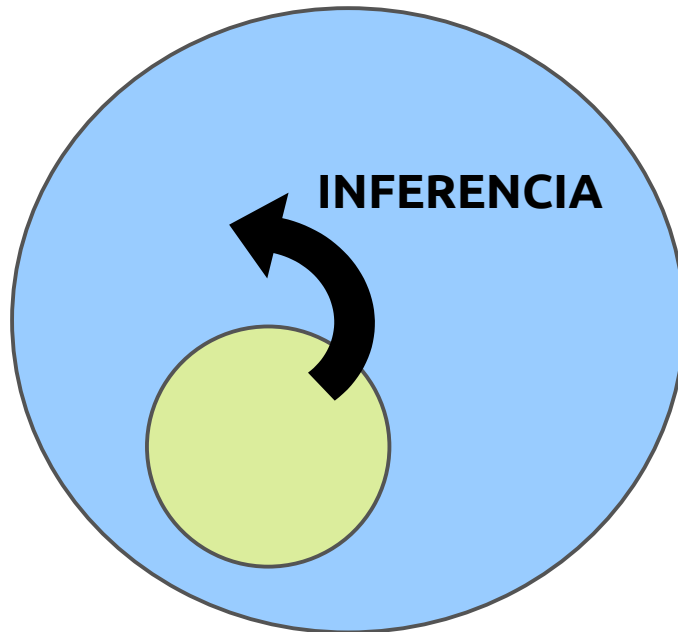
# El método científico



- Población: Conjunto **total de individuos u observaciones** sobre los que queremos sacar conclusiones.
- Muestra: **Subconjunto representativo** de la población, seleccionado para su estudio.



- Población: Conjunto **total de individuos u observaciones** sobre los que queremos sacar conclusiones.
- Muestra: **Subconjunto representativo** de la población, seleccionado para su estudio.



## Inferencia estadística

El proceso de usar los datos de una muestra para estimar o tomar decisiones sobre una población, reconociendo la incertidumbre del muestreo.

# DISEÑO EXPERIMENTAL VS DATOS OBSERVACIONALES

	Diseño experimental	Estudio observacional
Control	El investigador <b>manipula</b> una o más variables (tratamientos) y <b>controla</b> las condiciones.	El investigador <b>no manipula</b> nada; solo <b>observa y registra</b> lo que ocurre naturalmente.
Ejemplo		
Fuentes de variabilidad		

# DISEÑO EXPERIMENTAL VS DATOS OBSERVACIONALES

	Diseño experimental	Estudio observacional
Control	El investigador <b>manipula</b> una o más variables (tratamientos) y <b>controla</b> las condiciones.	El investigador <b>no manipula</b> nada; solo <b>observa y registra</b> lo que ocurre naturalmente.
Ejemplo	Ensayo clínico, experimento de laboratorio, cultivo en condiciones controladas.	Ecología de campo, estudios sociales, encuestas, registros médicos.
Fuentes de variabilidad		

# DISEÑO EXPERIMENTAL VS DATOS OBSERVACIONALES

	Diseño experimental	Estudio observacional
Control	El investigador <b>manipula</b> una o más variables (tratamientos) y <b>controla</b> las condiciones.	El investigador <b>no manipula</b> nada; solo <b>observa y registra</b> lo que ocurre naturalmente.
Ejemplo	Ensayo clínico, experimento de laboratorio, cultivo en condiciones controladas.	Ecología de campo, estudios sociales, encuestas, registros médicos.
Fuentes de variabilidad	Pueden <b>controlarse o aislarse</b> (aleatorización, réplicas, bloqueos).	Difíciles de controlar: el entorno, el comportamiento, la historia previa...

En experimentos, la estadística ayuda a diseñar los tratamientos, minimizar sesgos, y evaluar diferencias entre grupos. En estudios observacionales permite controlar confusores, modelar la variabilidad natural y extraer patrones sin control directo.

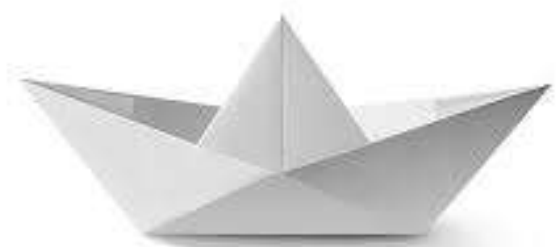
# **¿Qué es un modelo?**

# ¿Qué es un modelo?



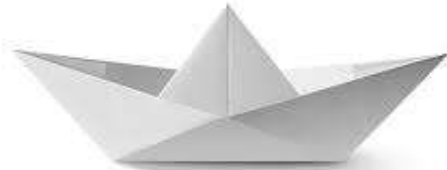


## ¿Qué es un modelo?



## ¿Qué es un modelo?

Un modelo es una **representación simplificada** de la realidad, elaborada con el fin de **explicar y/o predecir** ciertos aspectos de la misma



## ¿Qué es un modelo?

Un **modelo matemático** es la traducción de una hipótesis sobre la realidad a lenguaje matemático, de modo que pueda ponerse a prueba y contrastarse con datos.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

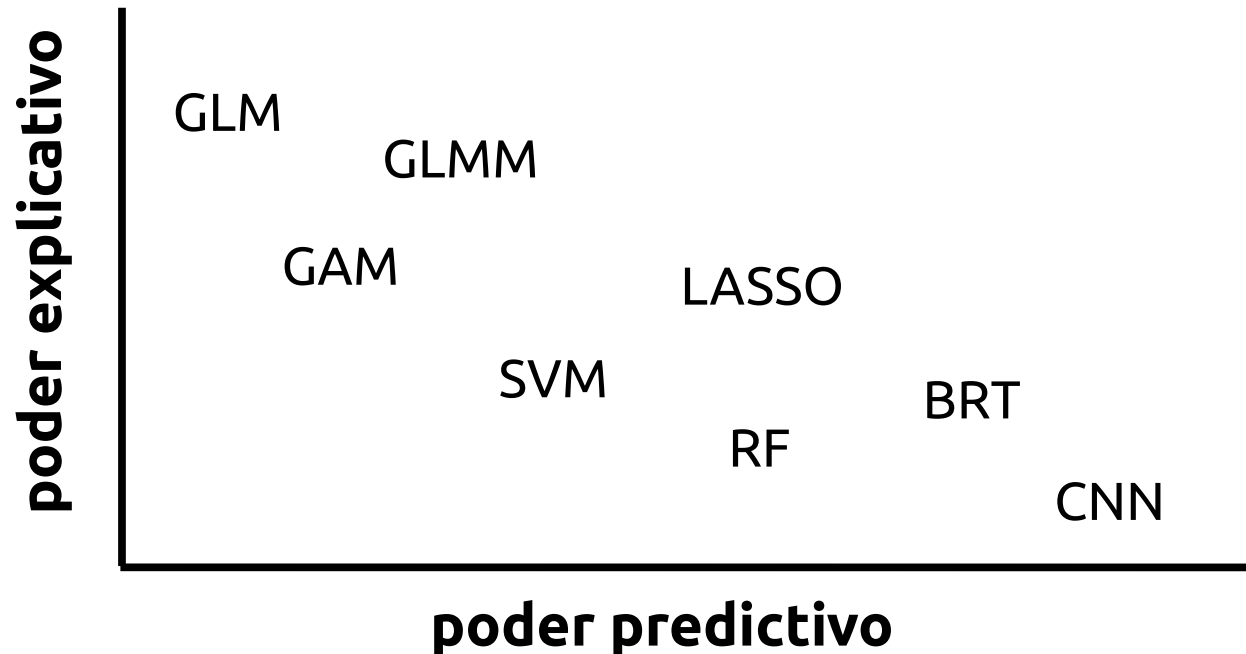


### Sobre explicaciones y predicciones..

Un modelo es una **representación simplificada** de la realidad, elaborada con el fin de **explicar y/o predecir** ciertos aspectos de la misma

## Sobre explicaciones y predicciones..

Un modelo es una **representación simplificada** de la realidad, elaborada con el fin de **explicar y/o predecir** ciertos aspectos de la misma



## Tipos de variables (clasificación clásica)

## Tipos de variables (clasificación clásica)

- Cuantitativas (se expresa mediante números, se puede operar)

Continuas (decimales)

Discretas (enteros)

## Tipos de variables (clasificación clásica)

- Cuantitativas (se expresa mediante números, se puede operar)

Continuas (decimales)

Discretas (enteros)

- Semicuantitativas u ordinales (variable cualitativa que se puede ordenar)



## Tipos de variables (clasificación clásica)

- Cuantitativas (se expresa mediante números, se puede operar)

Continuas (decimales)

Discretas (enteros)

- Semicuantitativas u ordinales (variable cualitativa que se puede ordenar)
- Cualitativas o nominales (describe un atributo o categoría)

## Tipos de variables (clasificación clásica)

- Cuantitativas (se expresa mediante números, se puede operar)

Continuas (decimales)

Discretas (enteros)

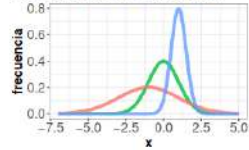
- Semicuantitativas u ordinales (variable cualitativa que se puede ordenar)
- Cualitativas o nominales (describe un atributo o categoría)

**¿Con qué variables trabajáis? ¿de qué tipo son?**

## Distribuciones de probabilidad

### Distribuciones continuas

#### Distribución Normal o Gaussiana

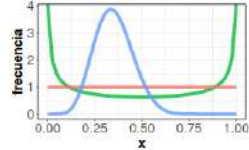


$X \sim \text{Normal}(\mu, \sigma)$   
 $X \sim \text{Normal}(\mu, \tau)$   
 $\tau = 1/\sigma$

**Dominio:**  
 $X \in (-\infty, \infty)$   
 $\mu \in (-\infty, \infty)$   
 $\sigma > 0$  (reales)

**R/NIMBLE:**  
`dnorm(mean, sd)`

#### Distribución Beta

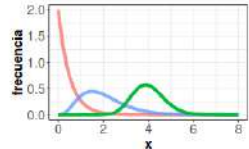


$X \sim \text{Beta}(\alpha, \beta)$   
 $X \sim \text{Beta}(\mu, \sigma)$   
 $\mu = \alpha / (\alpha + \beta)$

**Dominio:**  
 $X \in [0, 1]$   
 $\alpha > 0$  (real)  
 $\beta > 0$  (real)

**R/NIMBLE:**  
`dbeta(alpha, beta)`

#### Distribución Gamma



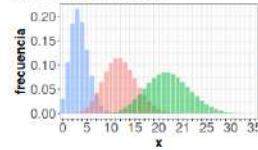
$X \sim \text{Gamma}(\alpha, \beta)$   
 $X \sim \text{Gamma}(\mu, \sigma)$   
 $\mu = \alpha / \beta$   
 $\sigma = \alpha / \beta^2$

**Dominio:**  
 $X \in (0, \infty)$   
 $\alpha > 0$  (real)  
 $\beta > 0$  (real)

**R/NIMBLE:**  
`dgamma(shape, rate)`

### Distribuciones discretas

#### Distribución de Poisson

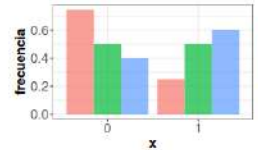


$X \sim \text{Poisson}(\lambda)$

**Dominio:**  
 $X \in \{0, \infty\}$  (naturales)  
 $\lambda \in (0, \infty)$  (reales)

**R/NIMBLE:**  
`dpois(lambda)`

#### Distribución de Bernoulli

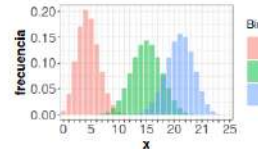


$X \sim \text{Bernoulli}(p)$

**Dominio:**  
 $X \in \{0, 1\}$   
 $p \in [0, 1]$

**R/NIMBLE:**  
`dbinom(1, prob)`  
`dbern(prob)`

#### Distribución Binomial



$X \sim \text{Binomial}(n, p)$

**Dominio:**  
 $X \in \{0, n\}$   
 $n \in [0, \infty)$  (natural)  
 $p \in [0, 1]$

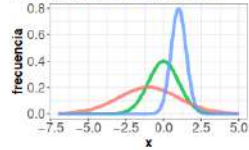
**R/NIMBLE:**  
`dbinom(n, size, prob)`  
`dbern(prob, size)`

- Van a ser nuestros modelos, nuestra abstracción matemática, la representación simplificada de la realidad!

## Distribuciones de probabilidad

### Distribuciones continuas

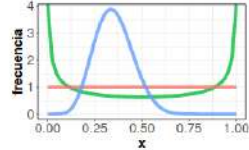
#### Distribución Normal o Gaussiana



**Dominio:**  
 $X \in (-\infty, \infty)$   
 $\mu \in (-\infty, \infty)$   
 $\sigma > 0$  (reales)

**R/NIMBLE:**  
`dnorm(mean, sd)`

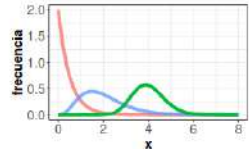
#### Distribución Beta



**Dominio:**  
 $X \in [0, 1]$   
 $\alpha > 0$  (real)  
 $\beta > 0$  (real)

**R/NIMBLE:**  
`dbeta(alpha, beta)`

#### Distribución Gamma

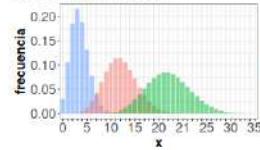


**Dominio:**  
 $X \in (0, \infty)$   
 $\alpha > 0$  (real)  
 $\beta > 0$  (real)

**R/NIMBLE:**  
`dgamma(shape, rate)`

### Distribuciones discretas

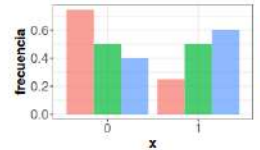
#### Distribución de Poisson



**Dominio:**  
 $X \in \{0, \infty\}$  (naturales)  
 $\lambda \in (0, \infty)$  (reales)

**R/NIMBLE:**  
`dpois(lambda)`

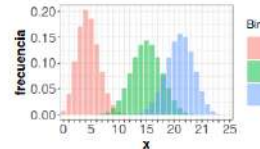
#### Distribución de Bernoulli



**Dominio:**  
 $X \in \{0, 1\}$   
 $p \in [0, 1]$

**R/NIMBLE:**  
`dbinom(1, prob)`  
`dbern(prob)`

#### Distribución Binomial



**Dominio:**  
 $X \in \{0, n\}$   
 $n \in [0, \infty)$  (natural)  
 $p \in [0, 1]$

**R/NIMBLE:**  
`dbinom(n, size, prob)`  
`dbern(prob, size)`

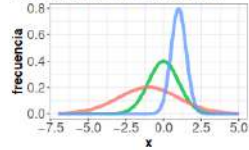
- Van a ser nuestros modelos, nuestra abstracción matemática, la representación simplificada de la realidad!

- Cada una tiene unas características diferentes que hacen que sean más aptas para uno u otro tipos de datos.

## Distribuciones de probabilidad

### Distribuciones continuas

#### Distribución Normal o Gaussiana

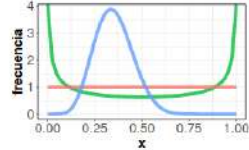


$X \sim \text{Normal}(\mu, \sigma)$   
 $X \sim \text{Normal}(\mu, \tau)$   
 $\tau = 1/\sigma$

**Domínio:**  
 $X \in (-\infty, \infty)$   
 $\mu \in (-\infty, \infty)$   
 $\sigma > 0$  (reales)

**R/NIMBLE:**  
`dnorm(mean, sd)`

#### Distribución Beta

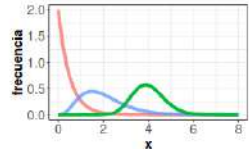


$X \sim \text{Beta}(\alpha, \beta)$   
 $X \sim \text{Beta}(\mu, \sigma)$   
 $\mu = \alpha / (\alpha + \beta)$   
 $\sigma = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

**Domínio:**  
 $X \in [0, 1]$   
 $\alpha > 0$  (real)  
 $\beta > 0$  (real)

**R/NIMBLE:**  
`dbeta(alpha, beta)`

#### Distribución Gamma



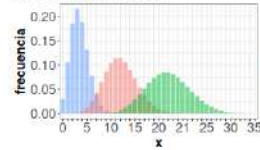
$X \sim \text{Gamma}(\alpha, \beta)$   
 $X \sim \text{Gamma}(\mu, \sigma)$   
 $\mu = \alpha/\beta$   
 $\sigma = \alpha/\beta^2$

**Domínio:**  
 $X \in (0, \infty)$   
 $\alpha > 0$  (real)  
 $\beta > 0$  (real)

**R/NIMBLE:**  
`dgamma(shape, rate)`

### Distribuciones discretas

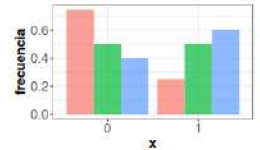
#### Distribución de Poisson



$X \sim \text{Poisson}(\lambda)$   
 $X \in \{0, \infty\}$  (naturales)  
 $\lambda \in (0, \infty)$  (reales)

**R/NIMBLE:**  
`dpois(lambda)`

#### Distribución de Bernoulli

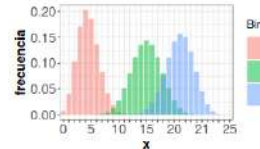


$X \sim \text{Bernoulli}(p)$

**Domínio:**  
 $X \in \{0, 1\}$   
 $p \in [0, 1]$

**R/NIMBLE:**  
`dbinom(1, prob)`  
`dbern(prob)`

#### Distribución Binomial



$X \sim \text{Binomial}(n, p)$

**Domínio:**  
 $X \in \{0, n\}$   
 $n \in [0, \infty)$  (natural)  
 $p \in [0, 1]$

**R/NIMBLE:**  
`dbinom(n, prob)`  
`dbinom(n, prob, size)`

- Van a ser nuestros modelos, nuestra abstracción matemática, la representación simplificada de la realidad!

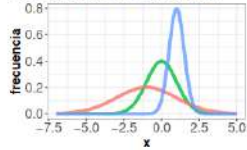
- Cada una tiene unas características diferentes que hacen que sean más aptas para uno u otro tipos de datos.

- Es necesario conocer las más importantes y sus principales características.

## Distribuciones de probabilidad

### Distribuciones continuas

#### Distribución Normal o Gaussiana

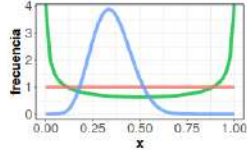


$X \sim \text{Normal}(\mu, \sigma)$   
 $X \sim \text{Normal}(\mu, \tau)$   
 $\tau = 1/\sigma$

**Domínio:**  
 $X \in (-\infty, \infty)$   
 $\mu \in (-\infty, \infty)$   
 $\sigma > 0$  (reales)

**R/NIMBLE:**  
`dnorm(mean, sd)`

#### Distribución Beta

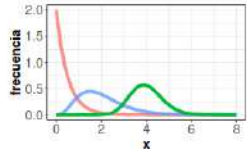


$X \sim \text{Beta}(\alpha, \beta)$   
 $X \sim \text{Beta}(\mu, \sigma)$   
 $\mu = \alpha / (\alpha + \beta)$   
 $\sigma = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

**Domínio:**  
 $X \in [0, 1]$   
 $\alpha > 0$  (real)  
 $\beta > 0$  (real)

**R/NIMBLE:**  
`dbeta(alpha, beta)`

#### Distribución Gamma



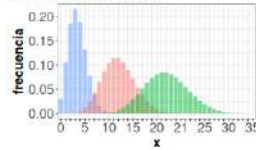
$X \sim \text{Gamma}(\alpha, \beta)$   
 $X \sim \text{Gamma}(\mu, \sigma)$   
 $\mu = \alpha / \beta$   
 $\sigma = \alpha / \beta^2$

**Domínio:**  
 $X \in (0, \infty)$   
 $\alpha > 0$  (real)  
 $\beta > 0$  (real)

**R/NIMBLE:**  
`dgamma(shape, rate)`

### Distribuciones discretas

#### Distribución de Poisson

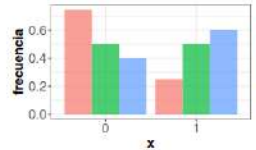


$X \sim \text{Poisson}(\lambda)$   
 $X \in \{0, \infty\}$  (naturales)  
 $\lambda \in (0, \infty)$  (reales)

**Domínio:**

**R/NIMBLE:**  
`dpois(lambda)`

#### Distribución de Bernoulli

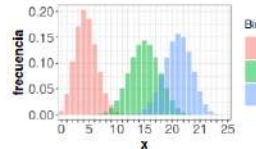


$X \sim \text{Bernoulli}(p)$

**Domínio:**  
 $X \in \{0, 1\}$   
 $p \in [0, 1]$

**R/NIMBLE:**  
`dbinom(1, prob)`  
`dbern(prob)`

#### Distribución Binomial



$X \sim \text{Binomial}(n, p)$

**Domínio:**  
 $X \in \{0, n\}$   
 $n \in [0, \infty)$  (natural)  
 $p \in [0, 1]$

**R/NIMBLE:**  
`dbinom(n, size, prob)`  
`dbinom(prob, size)`

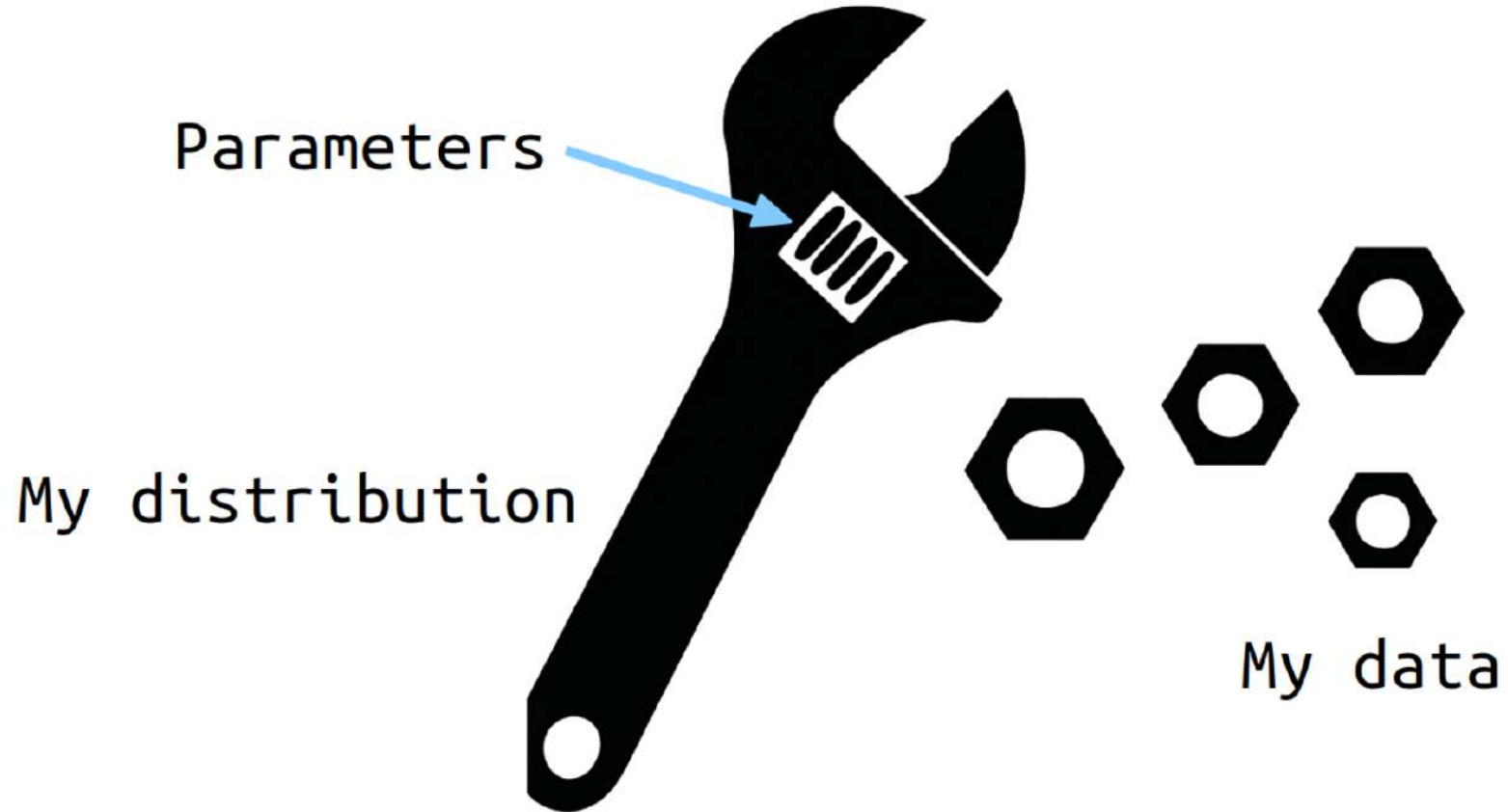
- Van a ser nuestros modelos, nuestra abstracción matemática, la representación simplificada de la realidad!

- Cada una tiene unas características diferentes que hacen que sean más aptas para uno u otro tipos de datos.

- Es necesario conocer las más importantes y sus principales características.

- Hay muchísimas, pero me gustaría que te aprendieses sólo tres (+1): la **Normal**, la de **Poisson** y la **Binomial** (+Uniforme).

## Distribuciones de probabilidad

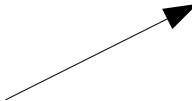


## Distribuciones de probabilidad: notación

$$X \sim \text{Name}(\text{parameter}_1, \text{parameter}_2)$$



## Distribuciones de probabilidad: notación

Mi variable   $X \sim \text{Name}(\text{parameter}_1, \text{parameter}_2)$

## Distribuciones de probabilidad: notación

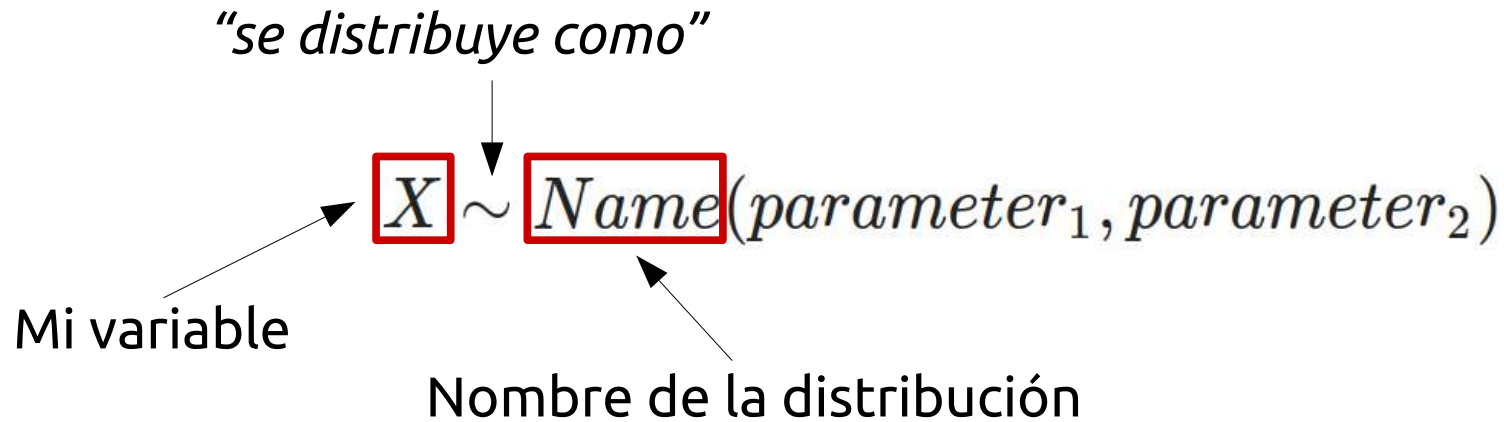
*"se distribuye como"*

$X$

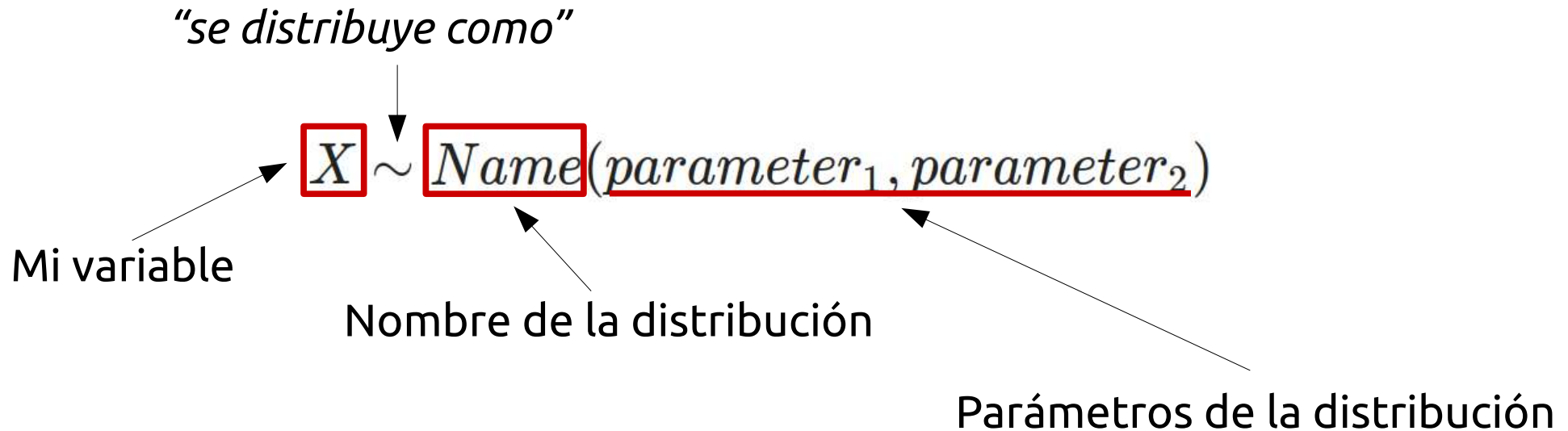
$\sim \text{Name}(\text{parameter}_1, \text{parameter}_2)$

Mi variable

## Distribuciones de probabilidad: notación



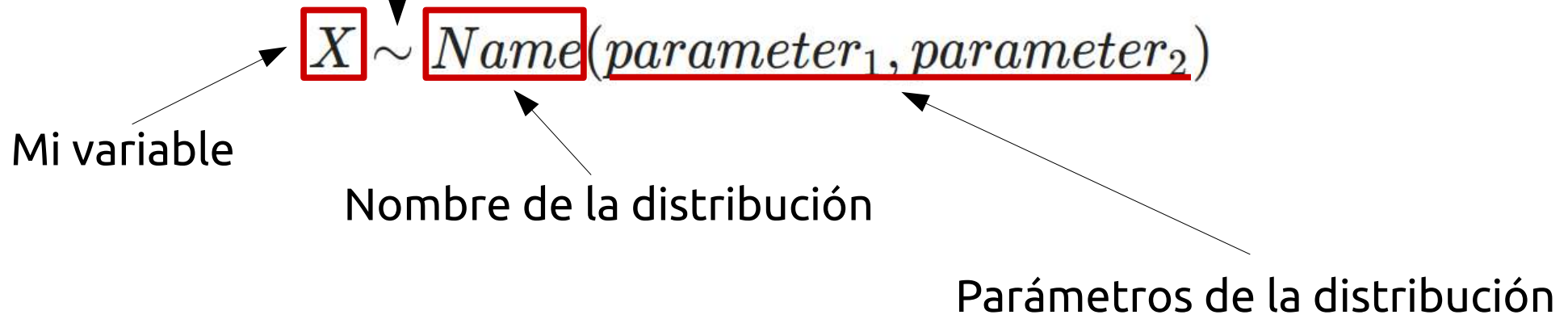
## Distribuciones de probabilidad: notación



## Distribuciones de probabilidad: notación

**OJO:** la misma distribución puede parametrizarse de diferentes formas

*"se distribuye como"*



## Distribuciones de probabilidad: notación

Temperatura media anual  $\sim$  Normal(media = 15, desviación estándar = 5)

## Distribuciones de probabilidad: notación

Temperatura media anual  $\sim$  Normal(media = 15, desviación estándar = 5)

Temperatura media anual  $\sim$  Normal( $\mu = 15$ ,  $\sigma = 5$ )

Temperatura media anual  $\sim N(15, 5)$

## Distribuciones de probabilidad: notación

Temperatura media anual ~ Normal(media = 15, desviación estándar = 5)

Temperatura media anual ~ Normal( $\mu = 15$ ,  $\sigma = 5$ )

Temperatura media anual ~ N(15, 5)

Standard deviation

$$X \sim \mathcal{N}(\mu, \sigma)$$



## Distribuciones de probabilidad: notación

Temperatura media anual ~ Normal(media = 15, desviación estándar = 5)

Temperatura media anual ~ Normal( $\mu = 15$ ,  $\sigma = 5$ )

Temperatura media anual ~ N(15, 5)

Standard deviation

$$X \sim \mathcal{N}(\mu, \sigma)$$

Variance

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

## Distribuciones de probabilidad: notación

Temperatura media anual  $\sim$  Normal(media = 15, desviación estándar = 5)

Temperatura media anual  $\sim$  Normal( $\mu = 15$ ,  $\sigma = 5$ )

Temperatura media anual  $\sim$  N(15, 5)

Standard deviation

$$X \sim \mathcal{N}(\mu, \sigma)$$

Variance

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Precision

$$X \sim \mathcal{N}(\mu, \tau)$$

$$\tau = \frac{1}{\sigma^2}$$

## Distribuciones de probabilidad: notación

Temperatura media anual ~ Normal(media = 15, desviación estándar = 5)

Temperatura media anual ~ Normal( $\mu = 15$ ,  $\sigma = 5$ )

Temperatura media anual ~ N(15, 5)

Standard deviation

$$X \sim \mathcal{N}(\mu, \sigma)$$

Variance

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Precision

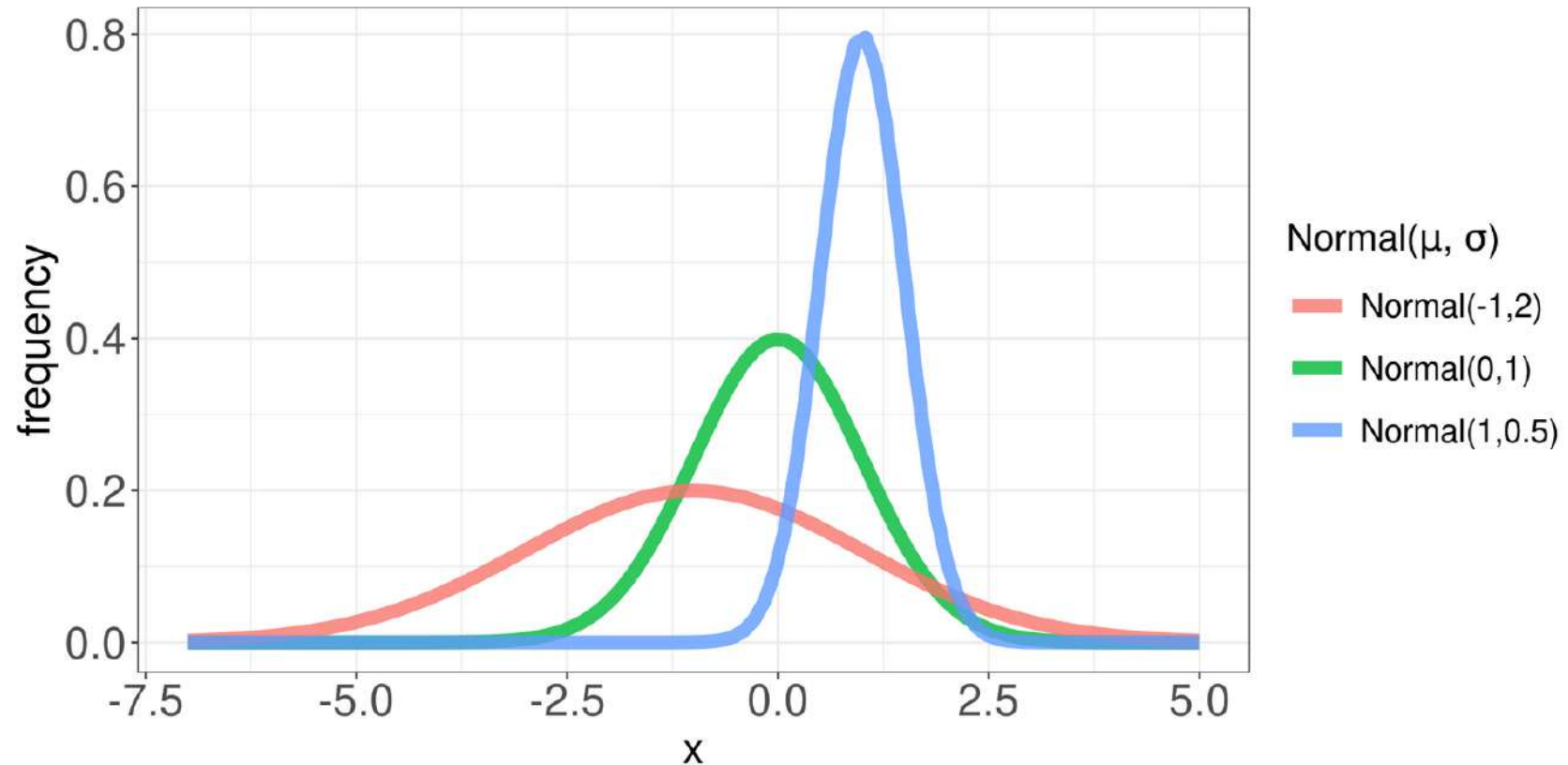
$$X \sim \mathcal{N}(\mu, \tau)$$

$$\tau = \frac{1}{\sigma^2}$$

$$X \sim \text{Normal}(\mu, \sigma)$$

**Continuas**

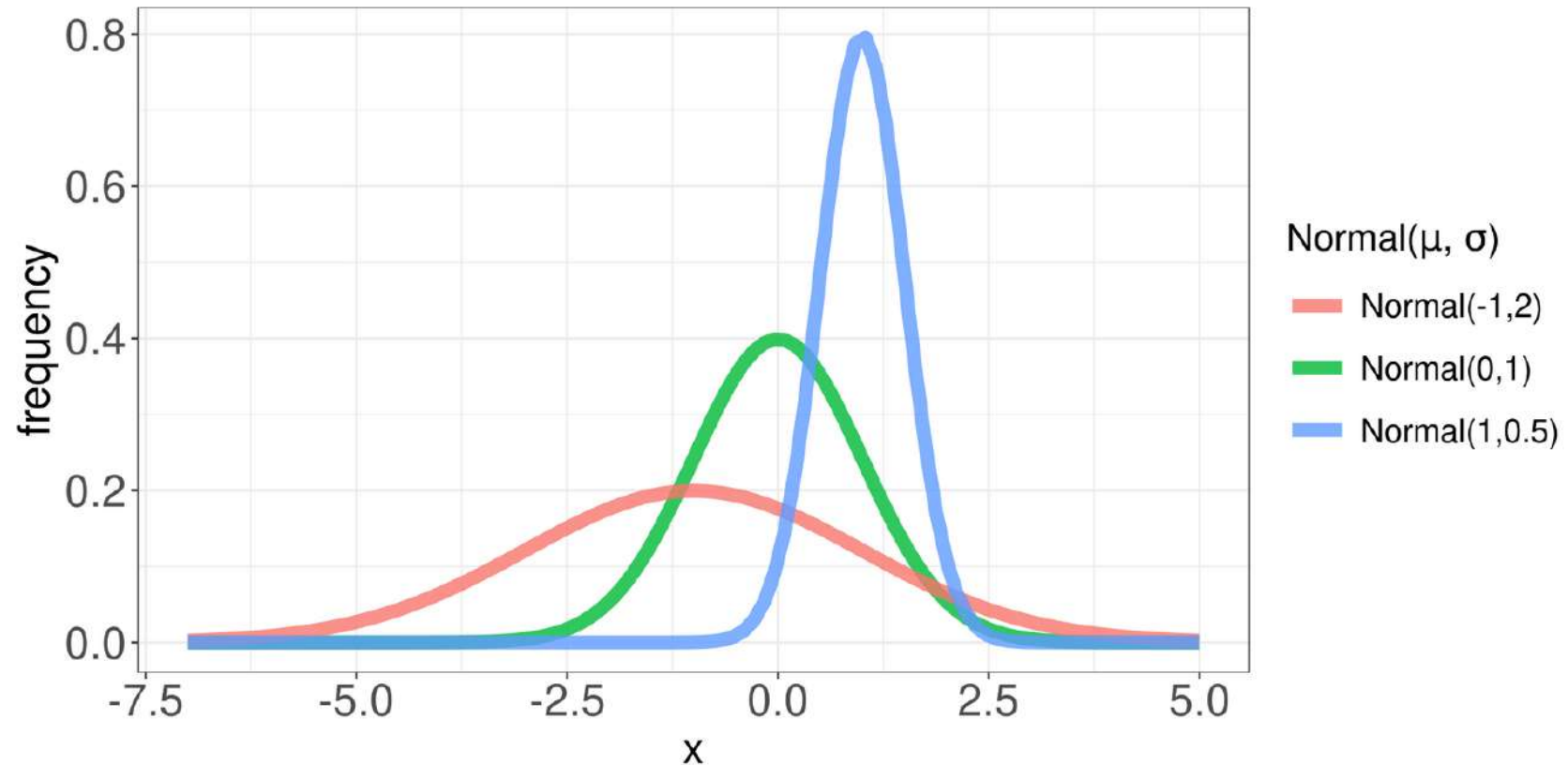
Support:  $X \in (-\infty, \infty); \mu \in (-\infty, \infty); \sigma > 0$  (real)



$$X \sim \text{Normal}(\mu, \sigma)$$

**Continuas**

Support:  $X \in (-\infty, \infty); \mu \in (-\infty, \infty); \sigma > 0$  (real)

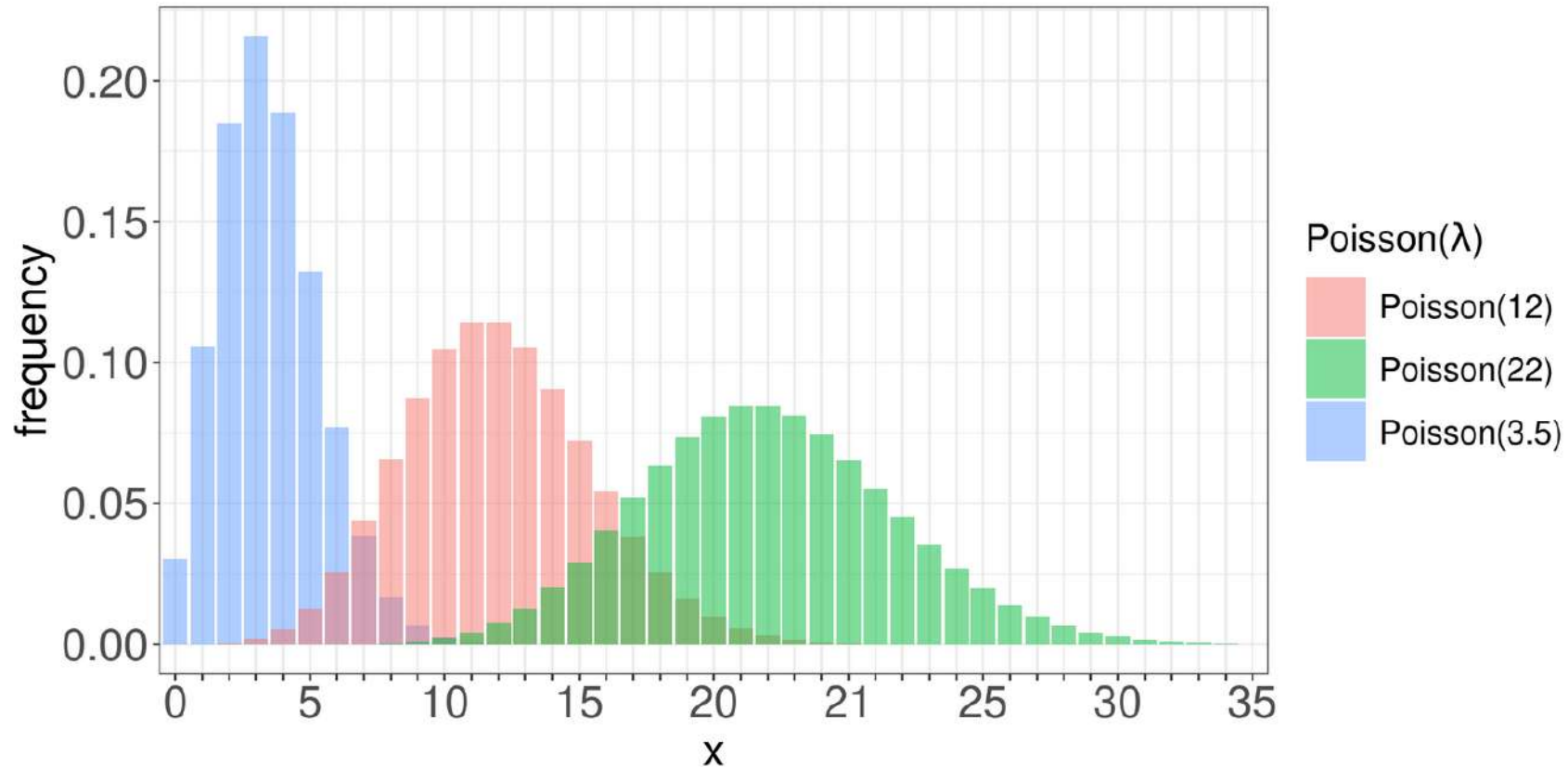


# VARIABLES Y DISTRIBUCIONES DE PROBABILIDAD

$$X \sim \text{Poisson}(\lambda)$$

**Conteos**

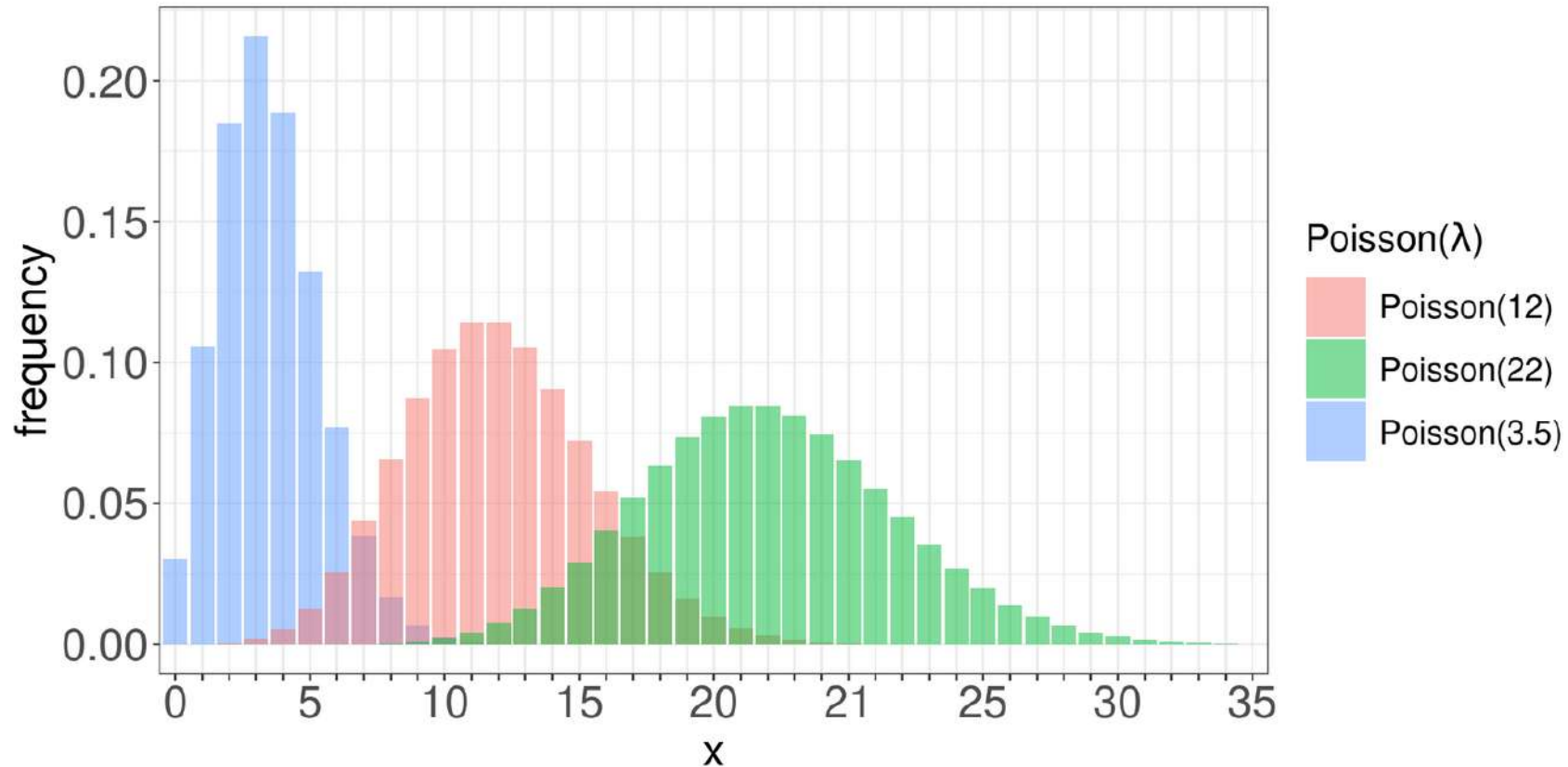
Support:  $X \in (0, \infty)$  (natural);  $\lambda \in (0, \infty)$  (real)



$$X \sim \text{Poisson}(\lambda)$$

**Conteos**

Support:  $X \in (0, \infty)$  (natural);  $\lambda \in (0, \infty)$  (real)

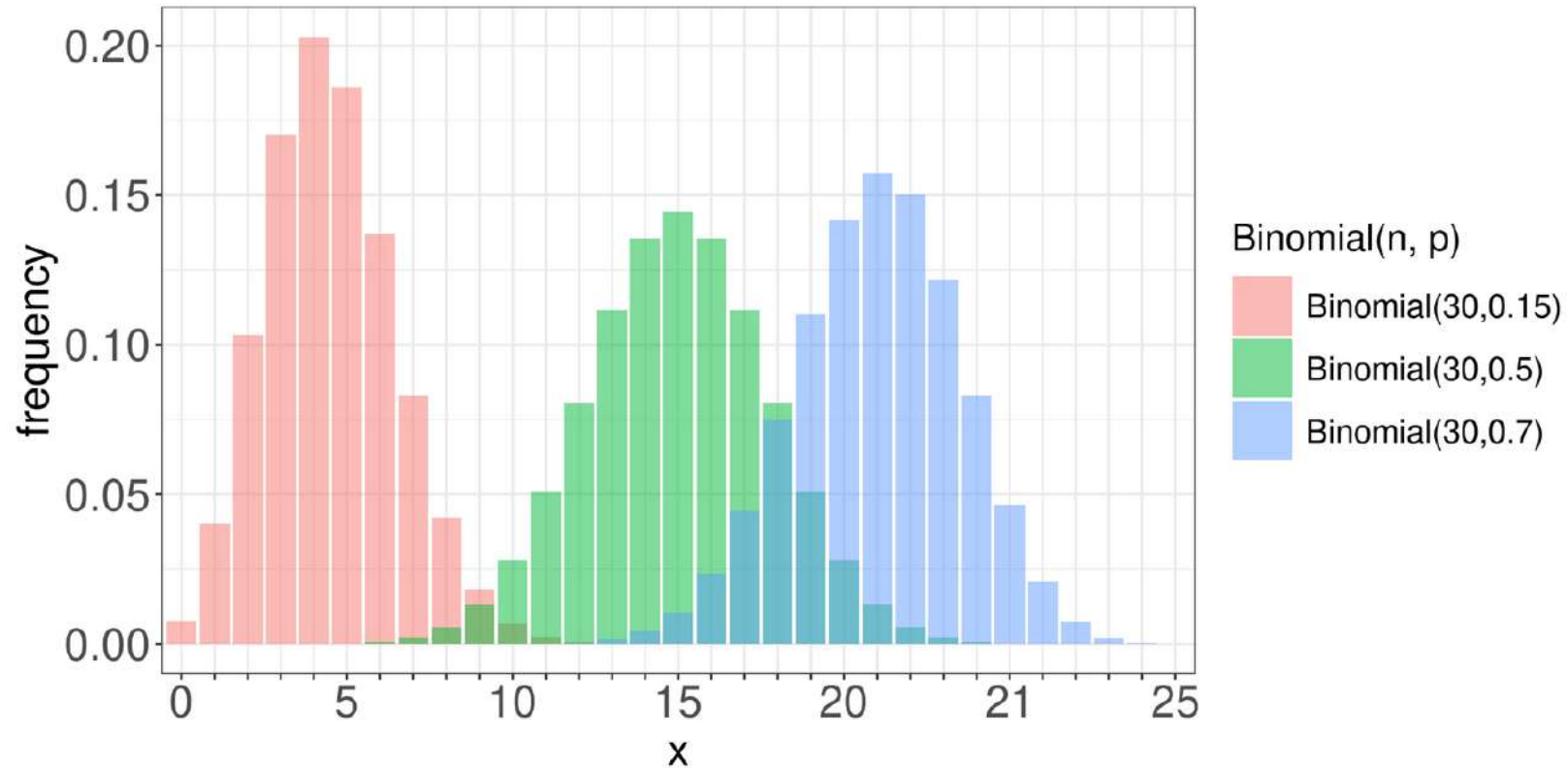


# VARIABLES Y DISTRIBUCIONES DE PROBABILIDAD

$$X \sim \text{Binomial}(n, p)$$

**Presencia/ausencia**

Support:  $X = \{0, 1\}; n \in (0, \infty)$  (natural)  $p \in (0, 1)$

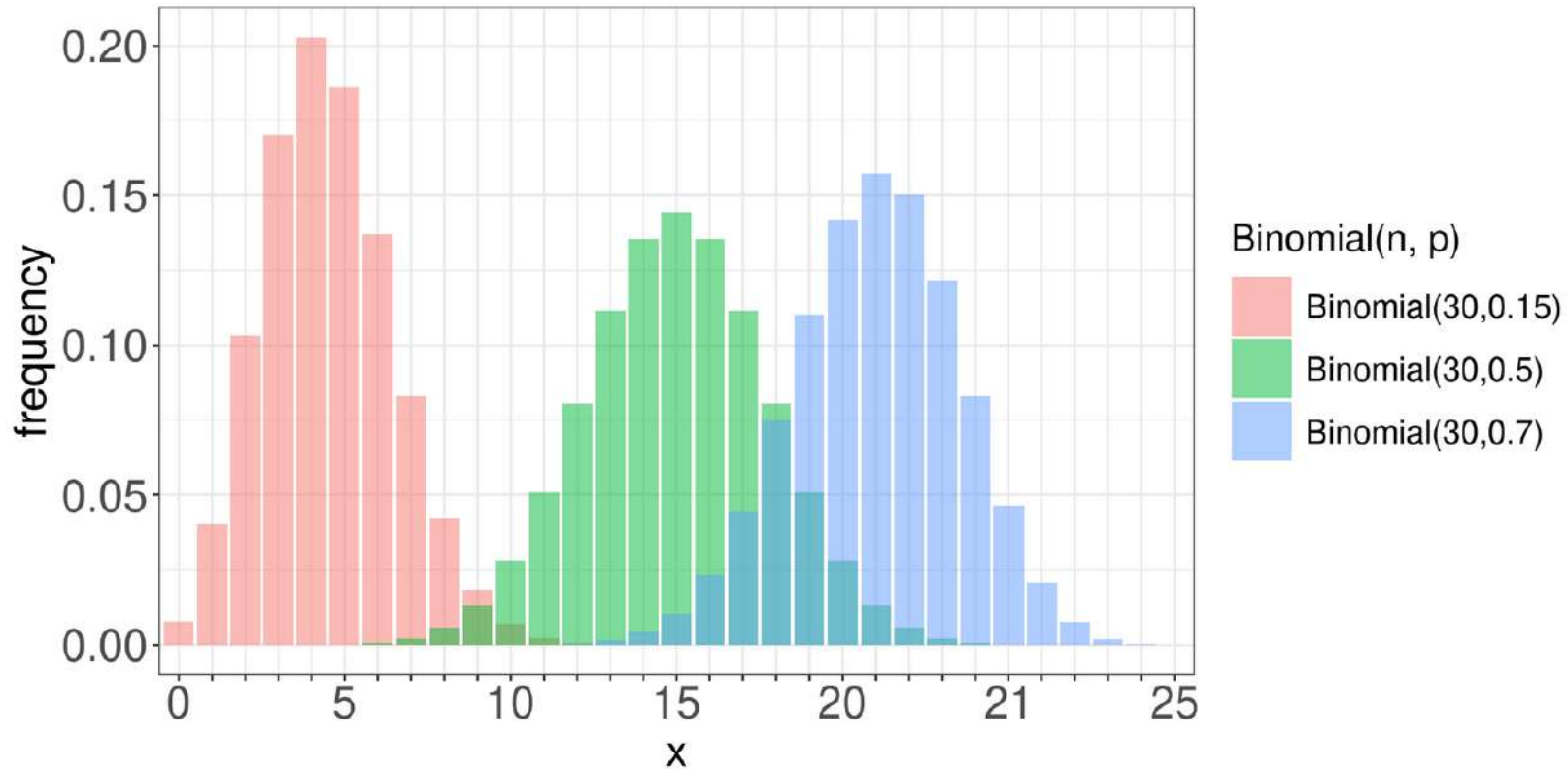




$$X \sim \text{Binomial}(n, p)$$

**Presencia/ausencia**

Support:  $X = \{0, 1\}; n \in (0, \infty)$  (natural)  $p \in (0, 1)$



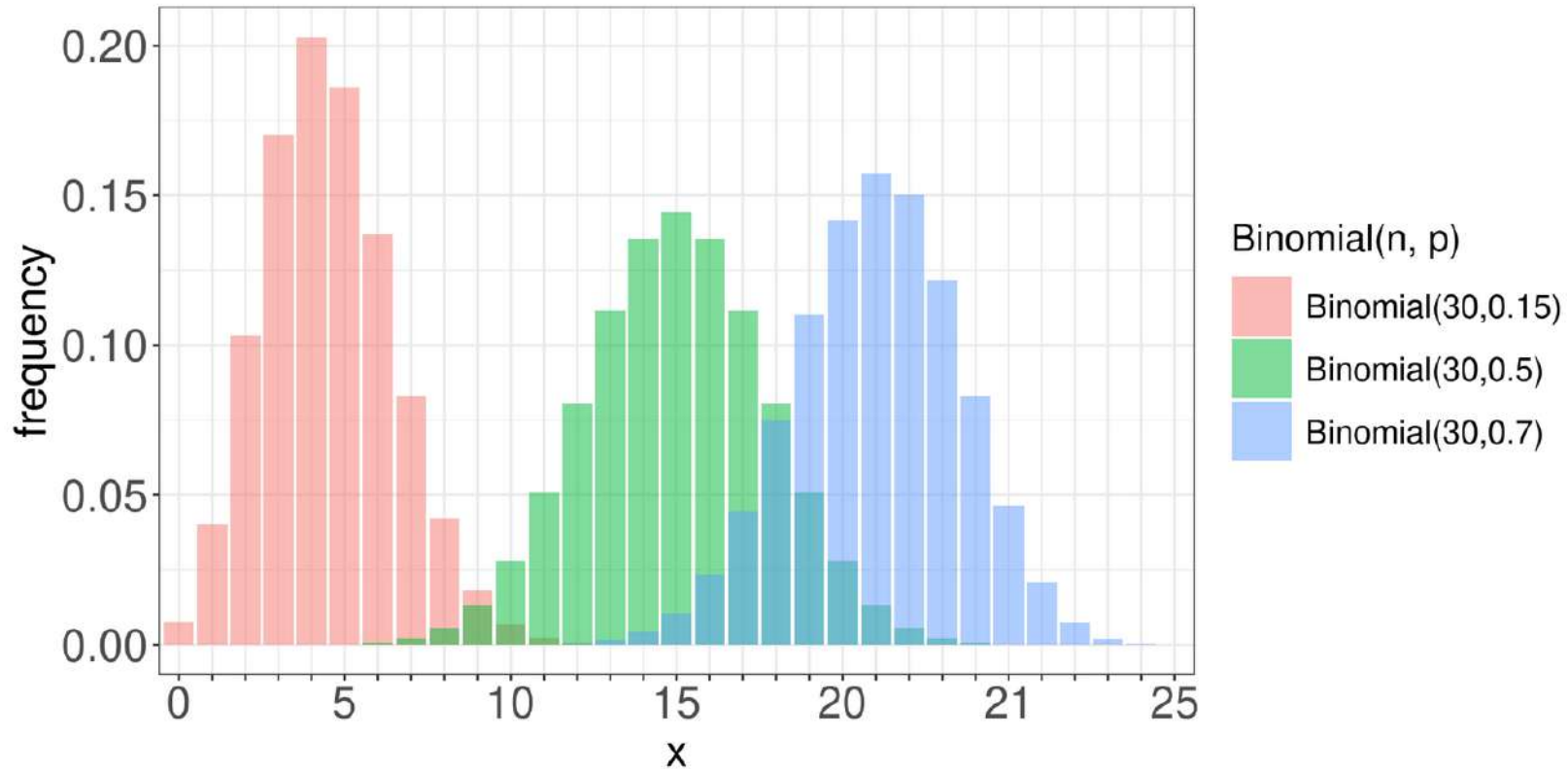
# VARIABLES Y DISTRIBUCIONES DE PROBABILIDAD

$$X \sim \text{Binomial}(n, p)$$

Support:  $X = \{0, 1\}; n \in (0, \infty)$  (natural)  $p \in (0, 1)$

**Presencia/ausencia**

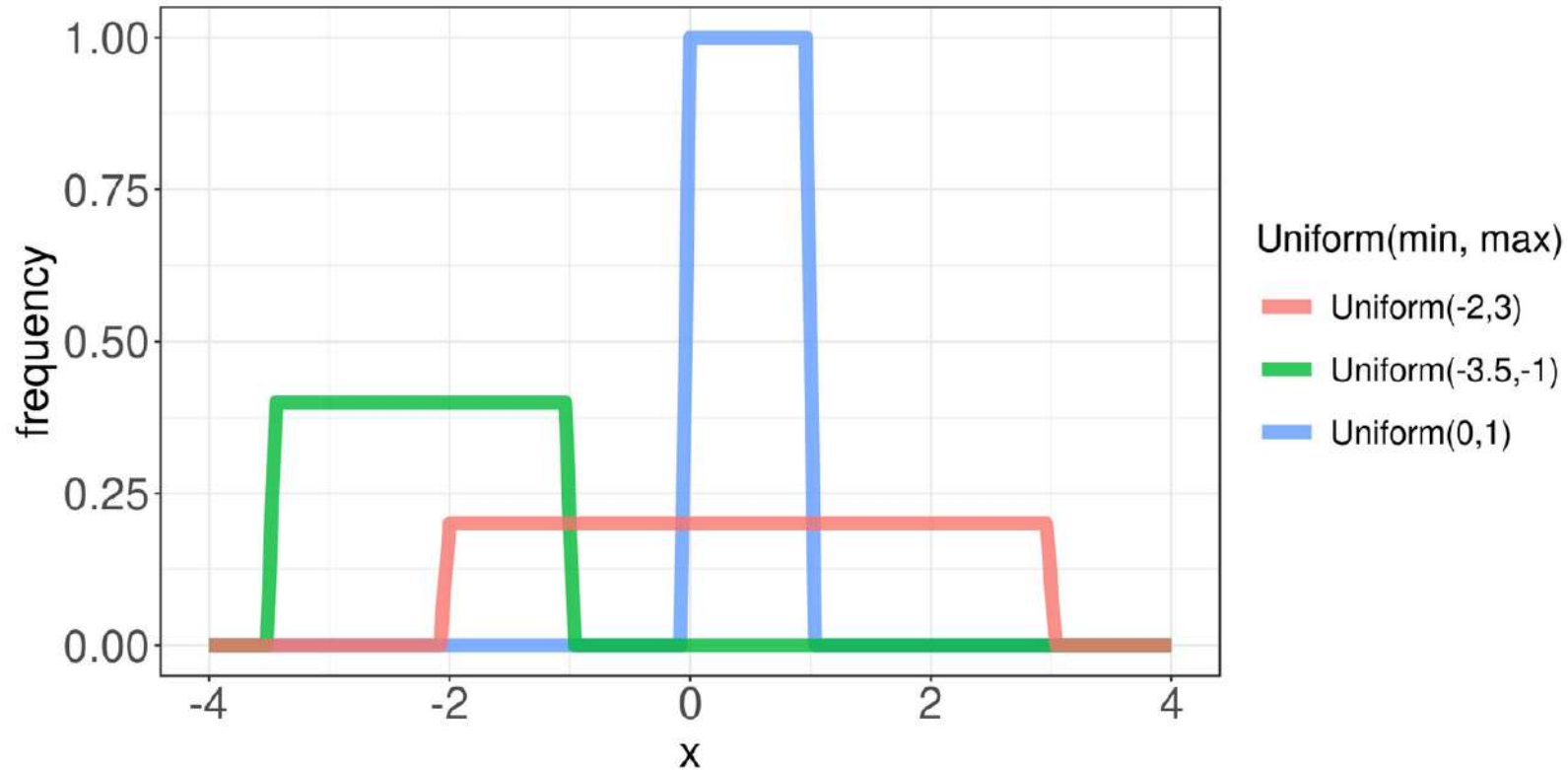
Bernoulli = Binomial(1,p)



# VARIABLES Y DISTRIBUCIONES DE PROBABILIDAD

$$X \sim \text{Uniform}(\min, \max)$$

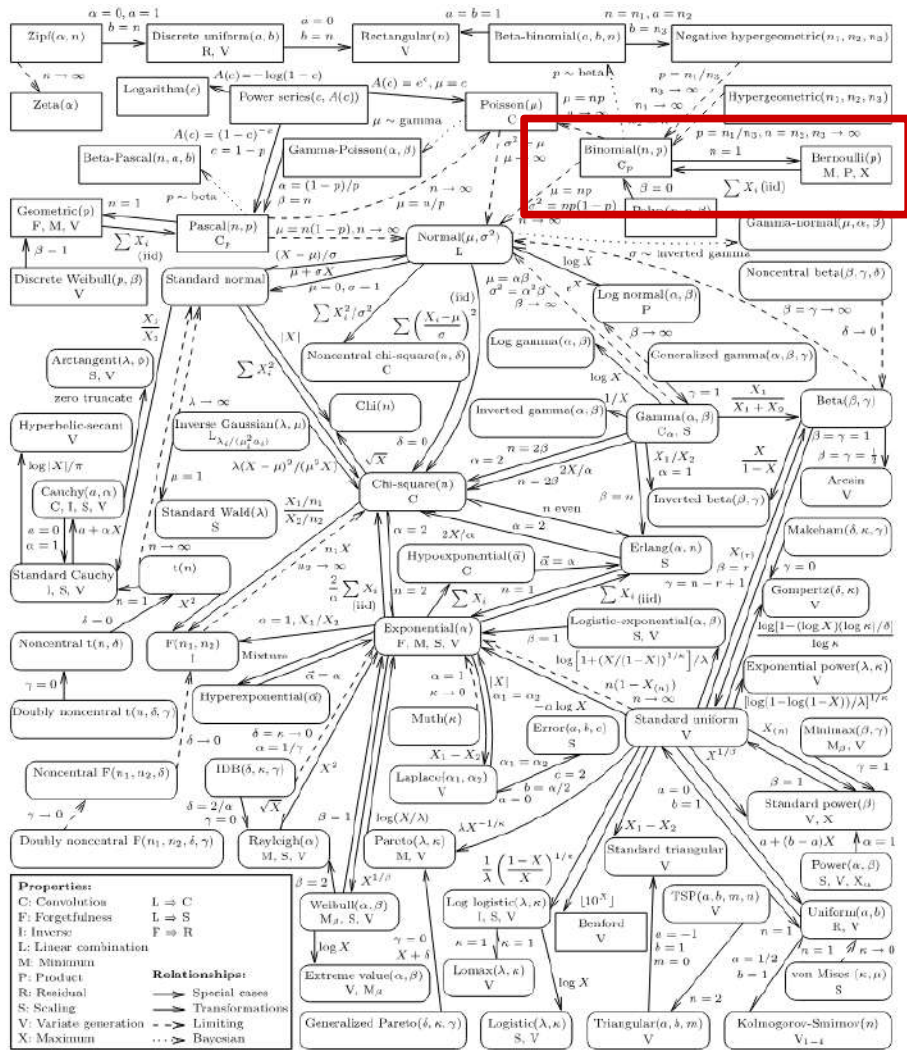
Support:  $X \in (-\infty, \infty)$ ;  $\min \in (-\infty, \infty)$  (real);  $\max \in (-\infty, \infty)$  (real)



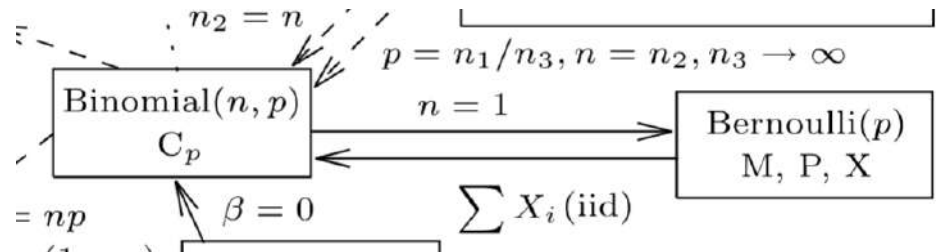
## Relación entre distribuciones de probabilidad univariadas



# VARIABLES Y DISTRIBUCIONES DE PROBABILIDAD



## Relación entre distribuciones de probabilidad univariadas



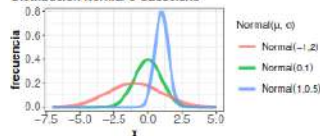
Leemis & McQueston, American Statistician (2008)



# VARIABLES Y DISTRIBUCIONES DE PROBABILIDAD

## Distribuciones continuas

### Distribución Normal o Gaussiana

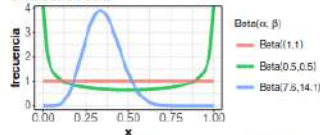


Normal( $\mu, \sigma$ )

$X \sim \text{Normal}(\mu, \sigma)$   
 $X \sim \text{Normal}(\mu, \tau)$   
 $\tau = 1/\sigma$

**R/NMBLE:**  
 dnorm(mean, sd)

### Distribución Beta

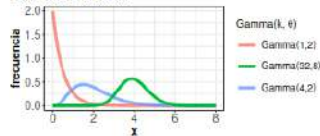


Beta( $\alpha, \beta$ )

$X \sim \text{Beta}(\alpha, \beta)$   
 $X \sim \text{Beta}(a, b)$   
 $a = \alpha / (\alpha + \beta)$   
 $b = \beta / (\alpha + \beta)$

**R/NMBLE:**  
 dbeta(alpha, beta)

### Distribución Gamma

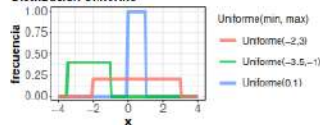


Gamma( $k, \theta$ )

$X \sim \text{Gamma}(\alpha, \beta)$   
 $X \sim \text{Gamma}(\mu, \sigma)$   
 $\mu = \alpha/\beta$   
 $\sigma = \alpha/\beta^2$

**R/NMBLE:**  
 dgamma(shape, rate)

### Distribución Uniforme



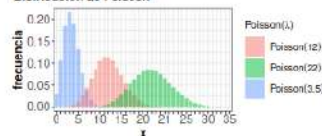
Uniforme(min, max)

$X \sim \text{Unif}(\min, \max)$   
 $\min \in (-\infty, \infty)$

**R/NMBLE:**  
 dunif(min, max)

## Distribuciones discretas

### Distribución de Poisson

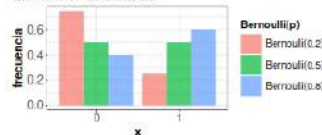


Poisson( $\lambda$ )

$X \sim \text{Poisson}(\lambda)$   
 $X \in \{0, 1, 2, \dots\}$  (naturales)  
 $\lambda \in \{0, \infty\}$  (reales)

**R/NMBLE:**  
 dpois(ambda)

### Distribución de Bernoulli

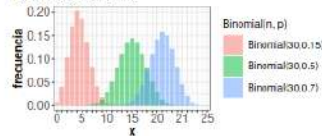


Bernoulli( $p$ )

$X \sim \text{Bernoulli}(p)$   
 $X \in \{0, 1\}$   
 $p \in [0, 1]$

**R/NMBLE:**  
 dbinom(1, prob)  
 dbern(prob)

### Distribución Binomial



Binomial( $n, p$ )

$X \sim \text{Binomial}(n, p)$   
 $X \in \{0, 1, 2, \dots, n\}$   
 $n \in \{0, \infty\}$  (natural)  
 $p \in [0, 1]$

**R/NMBLE:**  
 dbinom(size, prob)  
 dbern(prob, size)

## Principales funciones vínculo para GLM

Las funciones vínculo (*link functions*) sirven para enlazar los parámetros de las distribuciones de probabilidad con la ecuación general del modelo lineal, convirtiendo el modelo lineal *general* en *generalizado*. Los parámetros de algunas distribuciones no pueden tomar cualquier valor, por lo que se utilizan las funciones vínculo para adaptar los resultados de la ecuación del modelo lineal a los requerimientos del parámetro. En teoría (matemáticamente) es posible emplear cualquier función vínculo con cualquier distribución siempre que nuestros datos lo permitan. Sin embargo, en la práctica se suelen utilizar unas pocas funciones con cada una de las distribuciones de probabilidad. En la tabla se muestran las más comunes para las distribuciones estudiadas. Nótese que algunas distribuciones no se muestran con su parametrización más habitual.

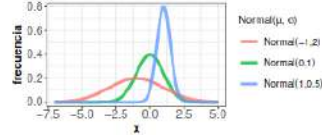
Distribución	Links	Fórmula	Inversa
Gaussiana( $\mu, \sigma$ )	<b>Identidad</b>	$\mu = \beta X$	$\mu = \beta X$
Beta( $\mu, \sigma$ )	<b>Logit</b> , probit, cloglog	$\log\left(\frac{\mu}{1-\mu}\right) = \beta X$	$\mu = \frac{e^{(\beta X)}}{1+e^{(\beta X)}}$
Gamma( $\mu, \sigma$ )	<b>Log</b>	$\log(\mu) = \beta X$	$\mu = e^{(\beta X)}$
Poisson( $\lambda$ )	<b>Log</b>	$\log(\lambda) = \beta X$	$\lambda = e^{(\beta X)}$
Bernoulli( $p$ )	<b>Logit</b> , probit, cloglog	$\log\left(\frac{p}{1-p}\right) = \beta X$	$p = \frac{e^{(\beta X)}}{1+e^{(\beta X)}}$
Binomial( $n, p$ )	<b>Logit</b> , probit, cloglog	$\log\left(\frac{p}{1-p}\right) = \beta X$	$p = \frac{e^{(\beta X)}}{1+e^{(\beta X)}}$



# VARIABLES Y DISTRIBUCIONES DE PROBABILIDAD

## Distribuciones continuas

### Distribución Normal o Gaussiana

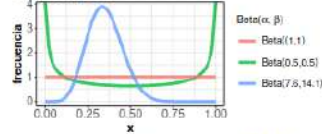


Normal( $\mu, \sigma$ )

Dominio:  
 $X \sim \text{Normal}(\mu, \sigma)$   
 $\mu \in (-\infty, \infty)$   
 $\sigma > 0$  (reales)

R/NMBLE:  
dnorm(mean, sd)

### Distribución Beta

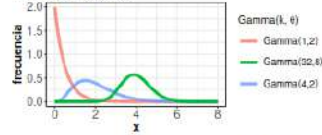


Beta( $\alpha, \beta$ )

Dominio:  
 $X \in [0, 1]$   
 $\alpha > 0$  (real)  
 $\beta > 0$  (real)

R/NMBLE:  
dbeta(alpha, beta)

### Distribución Gamma

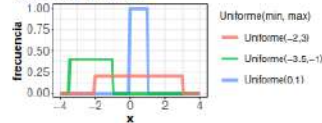


Gamma( $k, \theta$ )

Dominio:  
 $X \in (0, \infty)$   
 $\alpha > 0$  (real)  
 $\beta > 0$  (real)

R/NMBLE:  
dgamma(shape, rate)

### Distribución Uniforme



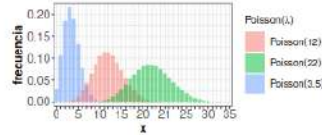
Uniforme(min, max)

Dominio:  
 $X \in (-\infty, \infty)$   
 $\min \in (-\infty, \infty)$  (real)

R/NMBLE:  
dunif(min, max)

## Distribuciones discretas

### Distribución de Poisson

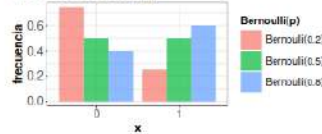


Poisson( $\lambda$ )

Dominio:  
 $X \sim \text{Poisson}(\lambda)$   
 $X \in \{0, \dots\}$  (naturales)  
 $\lambda \in \{0, \dots\}$  (reales)

R/NMBLE:  
dpois(lambda)

### Distribución de Bernoulli

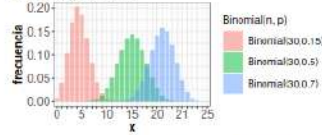


Bernoulli( $p$ )

Dominio:  
 $X \in \{0, 1\}$   
 $p \in [0, 1]$

R/NMBLE:  
dbinom(1, prob)

### Distribución Binomial



Binomial( $n, p$ )

Dominio:  
 $X \in \{0, \dots, n\}$   
 $n \in \{0, \dots\}$  (naturales)  
 $p \in [0, 1]$

R/NMBLE:  
dbinom(size, prob)

¿Con qué distribuciones de probabilidad habéis trabajado?

¿Qué otras distribuciones de probabilidad conocéis?

Matt Bognar website:

<https://homepage.divms.uiowa.edu/~mbognar/>



[https://jabiologo.github.io/web/tutorials/prob\\_dist\\_links\\_merged.pdf](https://jabiologo.github.io/web/tutorials/prob_dist_links_merged.pdf)

# Las simulaciones en el proceso de aprendizaje





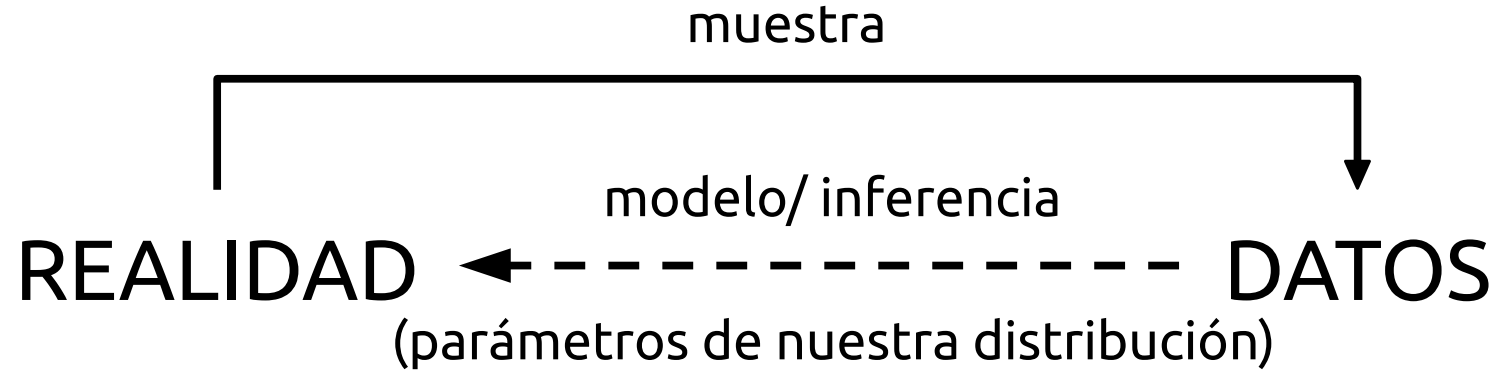
## Las simulaciones en el proceso de aprendizaje

REALIDAD

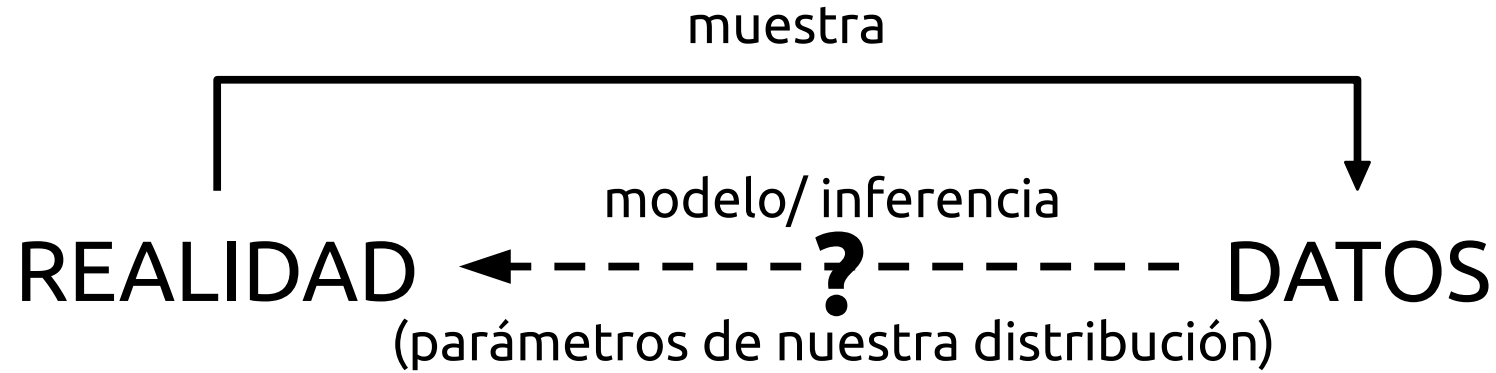
## Las simulaciones en el proceso de aprendizaje



## Las simulaciones en el proceso de aprendizaje



## Las simulaciones en el proceso de aprendizaje



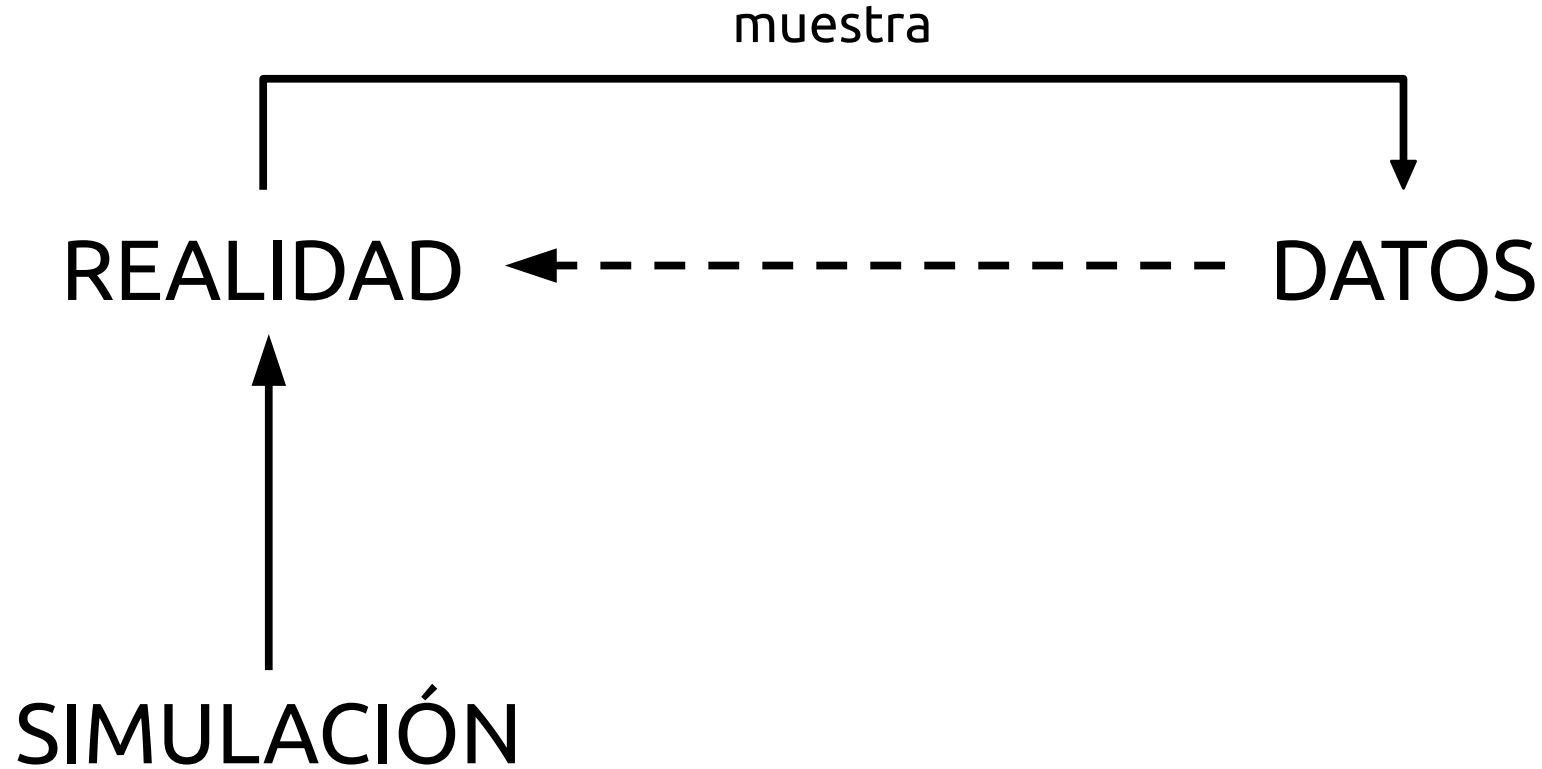
## Las simulaciones en el proceso de aprendizaje



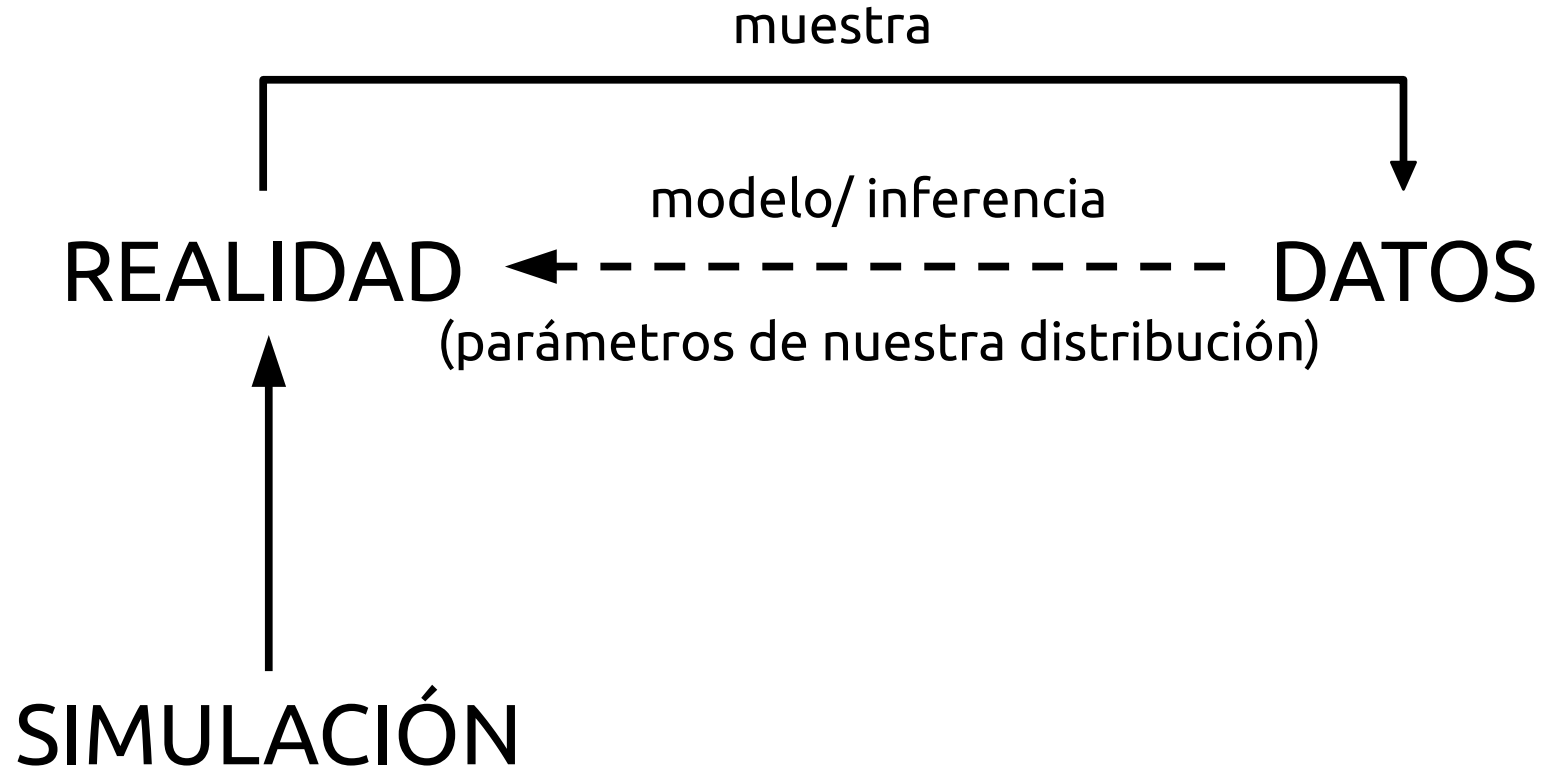
## Las simulaciones en el proceso de aprendizaje



## Las simulaciones en el proceso de aprendizaje

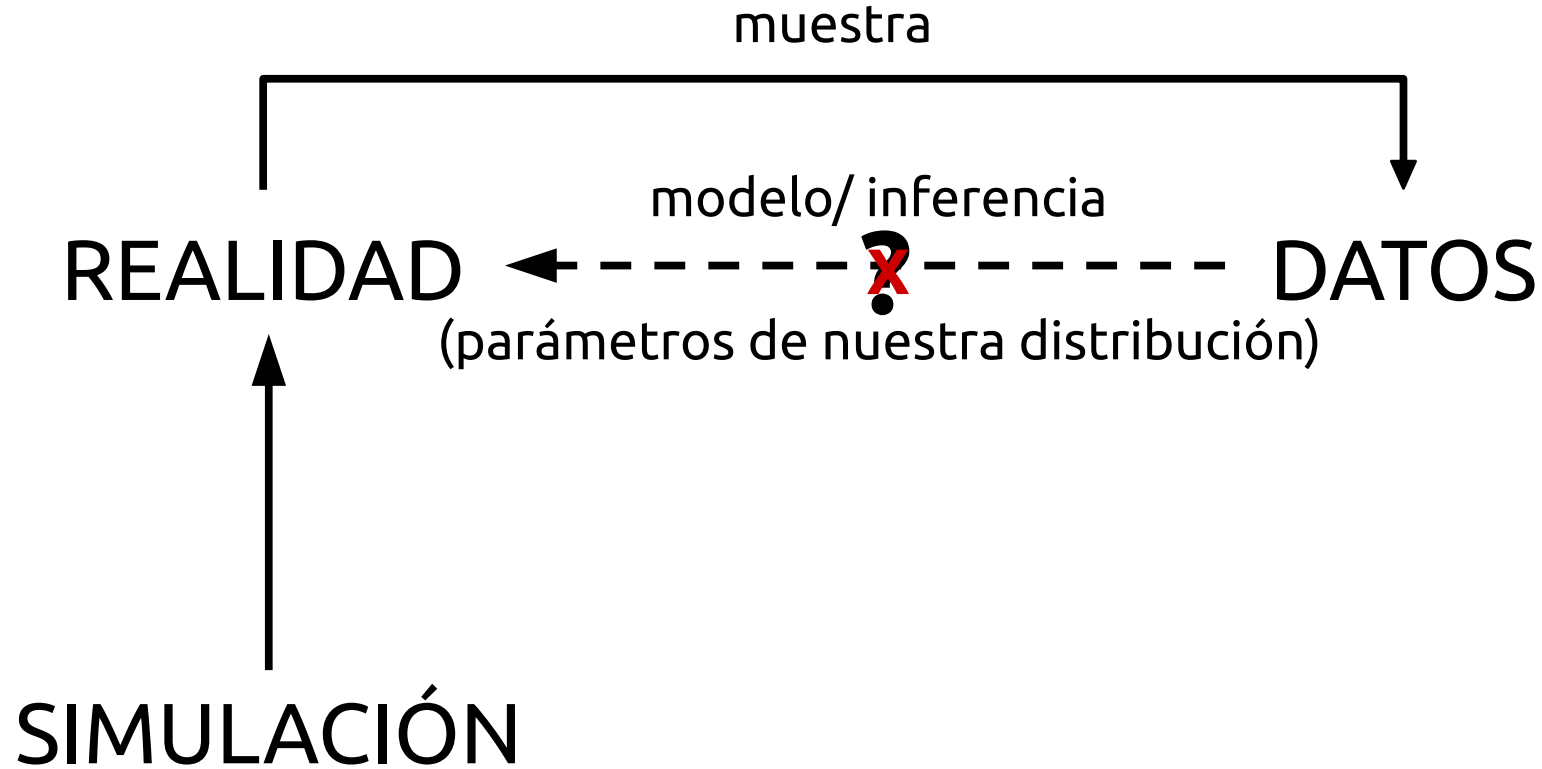


## Las simulaciones en el proceso de aprendizaje

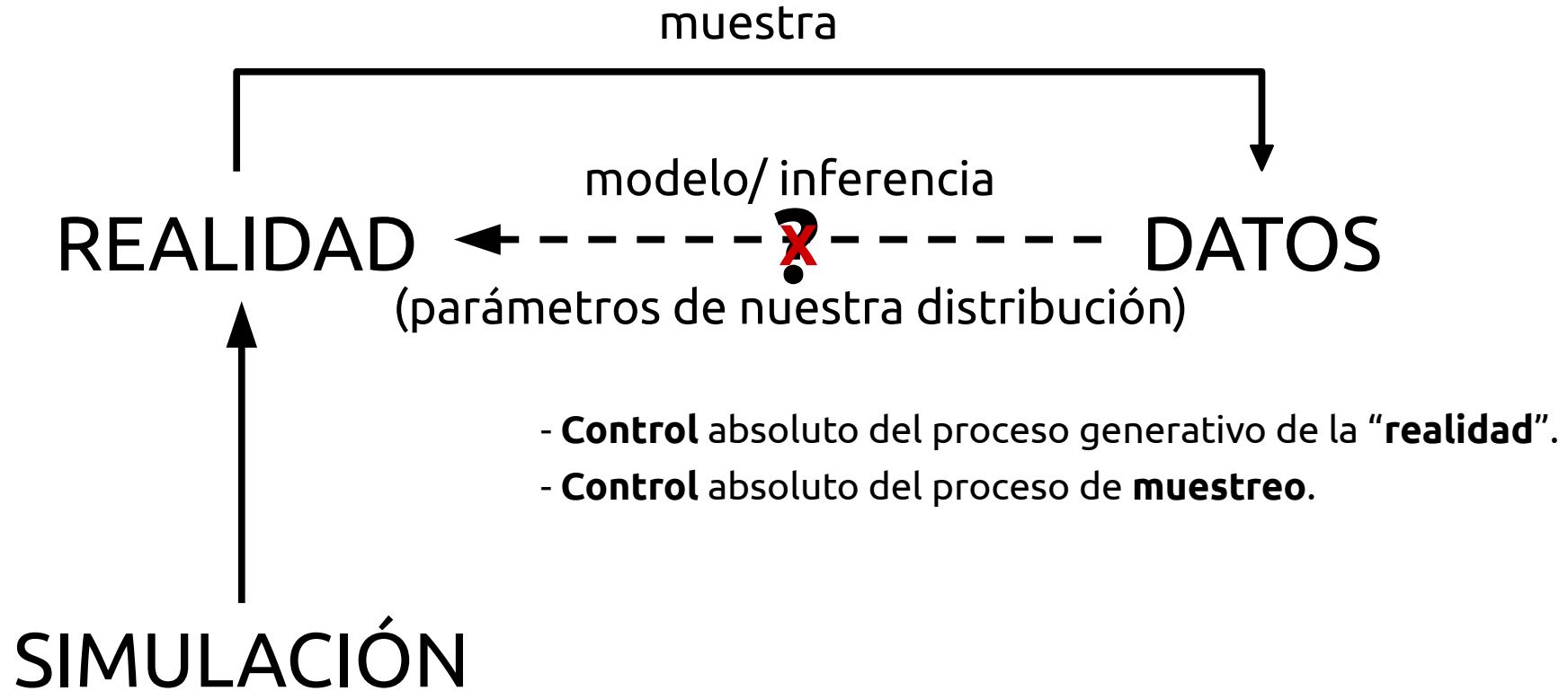




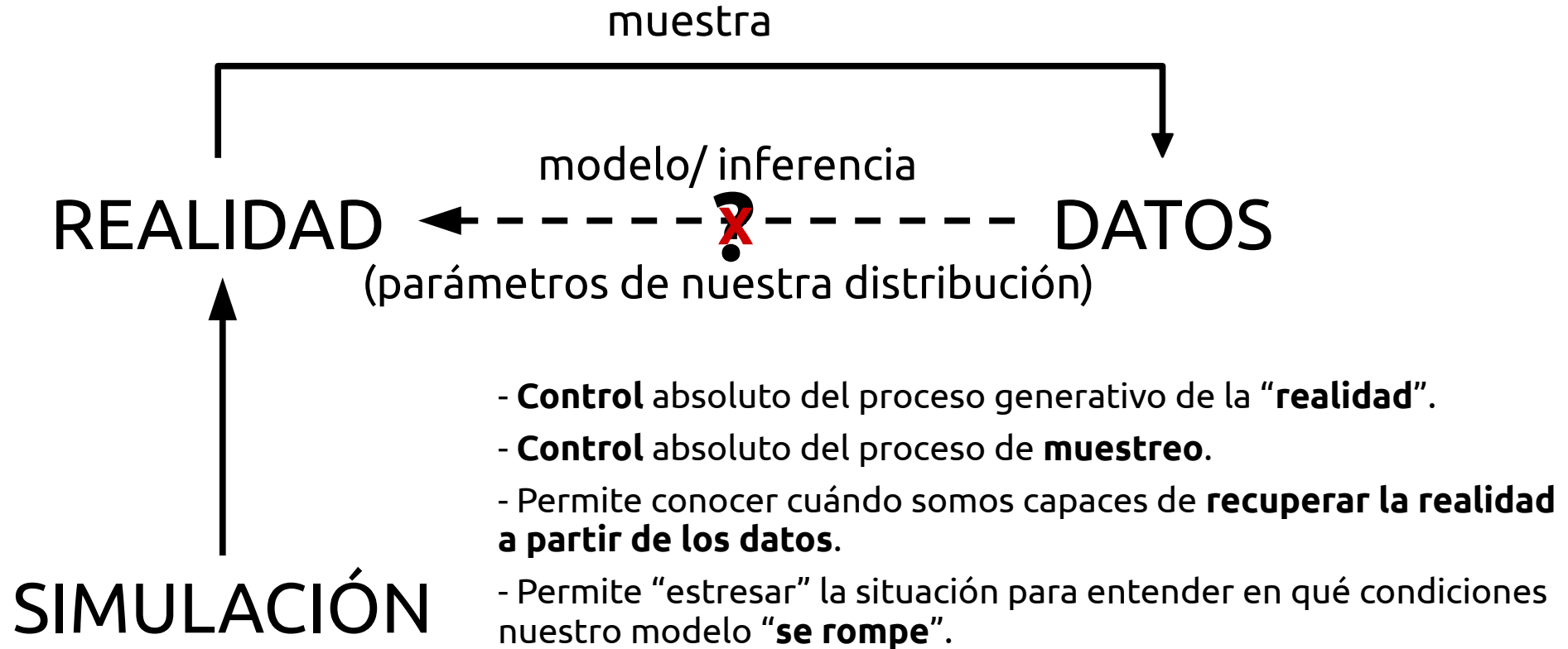
## Las simulaciones en el proceso de aprendizaje



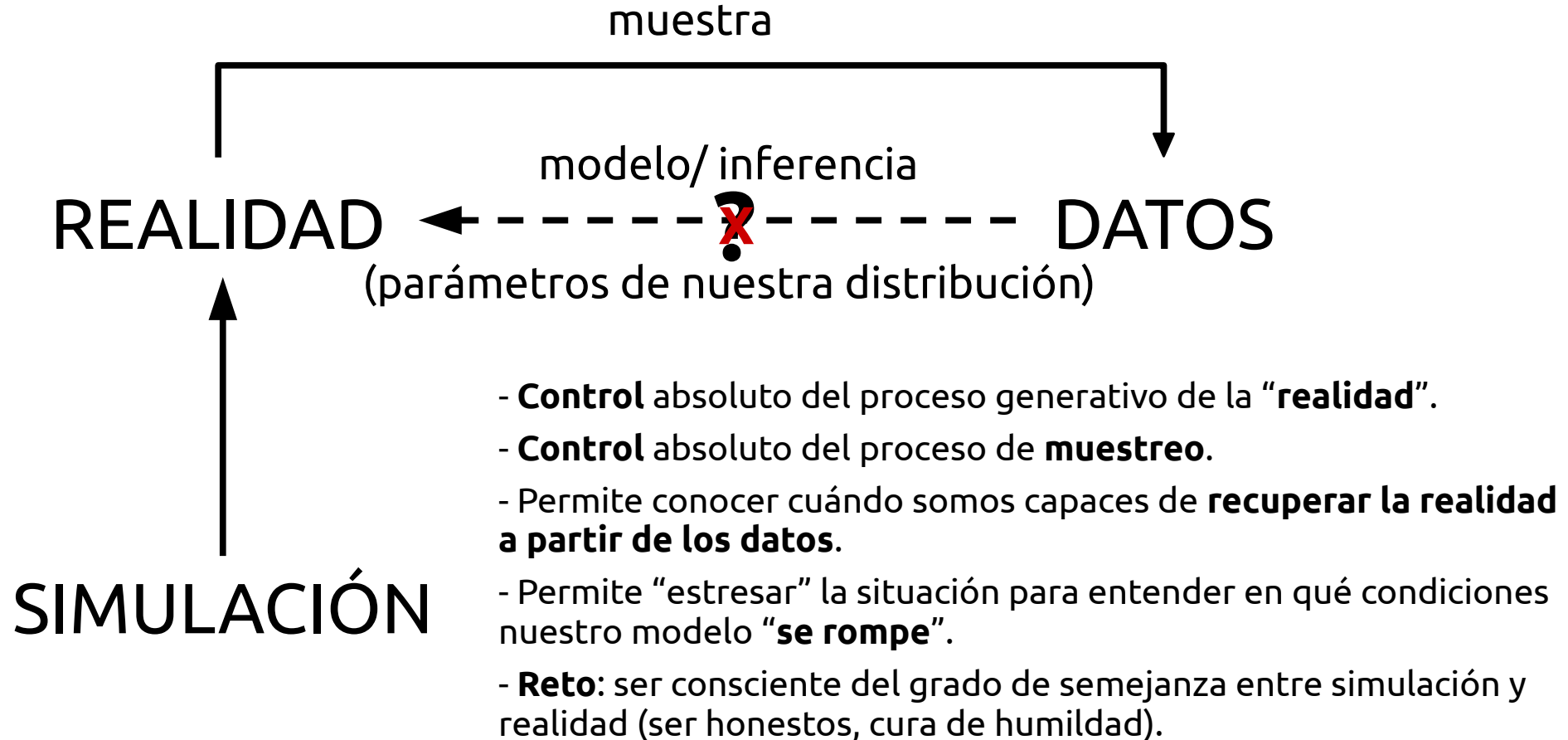
## Las simulaciones en el proceso de aprendizaje



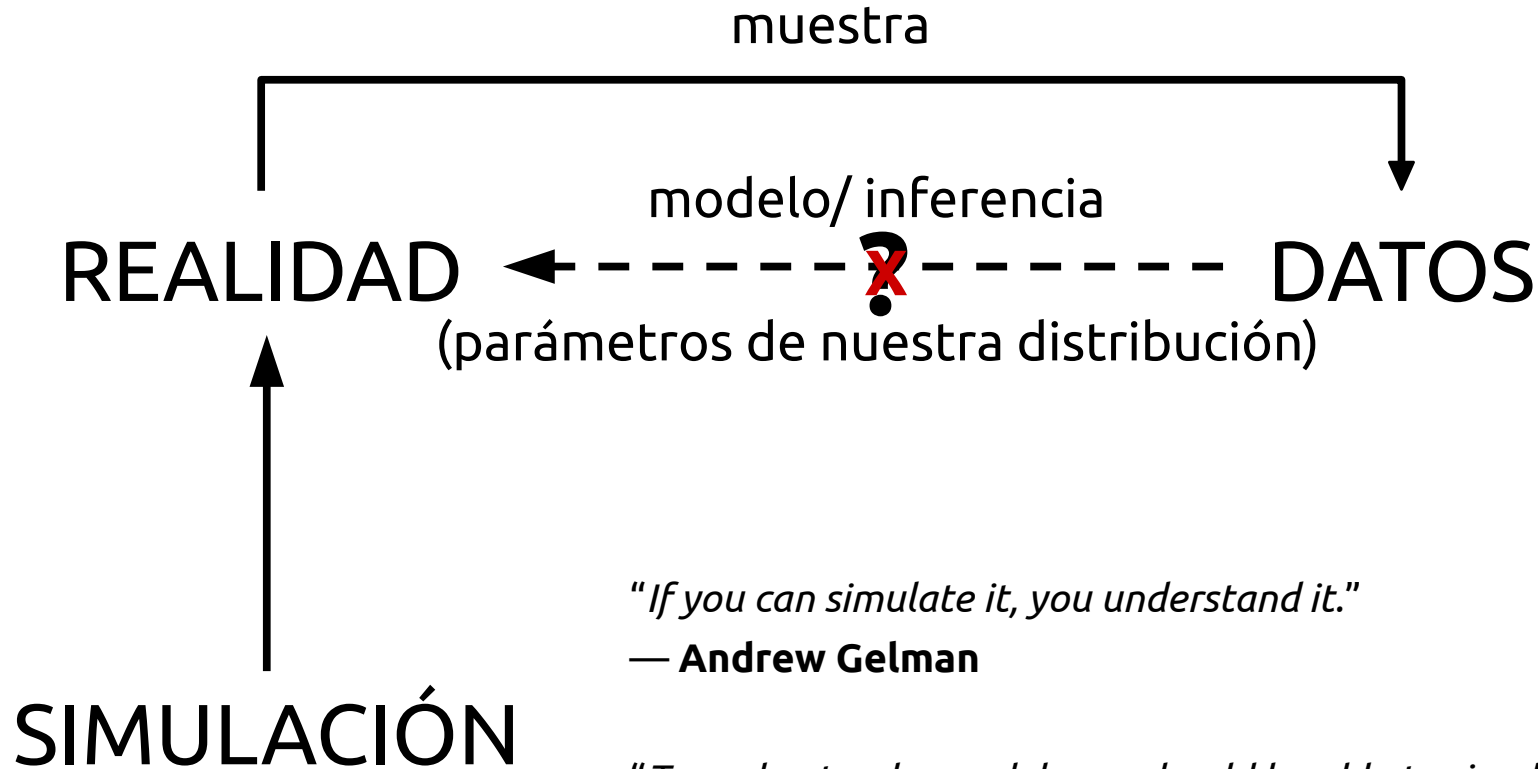
## Las simulaciones en el proceso de aprendizaje



## Las simulaciones en el proceso de aprendizaje



## Las simulaciones en el proceso de aprendizaje



*"If you can simulate it, you understand it."*

— **Andrew Gelman**

*"To understand a model, you should be able to simulate data from it."*

— **George Box**

# LA SIMULACIÓN COMO UN LABORATORIO

A screenshot of the RStudio application window. The title bar says 'RStudio'. The menu bar includes 'File', 'Edit', 'Code', 'View', 'Plots', 'Session', 'Build', 'Debug', 'Profile', 'Tools', and 'Help'. Below the menu bar is a toolbar with icons for file operations and a 'Go to file/function' search bar. The main interface is divided into four panes. The top-left pane is the 'Console', which shows the R startup message for version 4.5.0 (2025-04-11) on a Linux platform. The top-right pane is the 'Environment' pane, showing 'Project: (None)'. The bottom-left pane is the 'Files' pane, showing a file explorer view. The bottom-right pane is the 'Plots' pane, showing a plot area. The console text is as follows:

```
R version 4.5.0 (2025-04-11) -- "How About a Twenty-Six"  
Copyright (C) 2025 The R Foundation for Statistical Computing  
Platform: x86_64-pc-linux-gnu  
  
R es un software libre y viene sin GARANTIA ALGUNA.  
Usted puede redistribuirlo bajo ciertas circunstancias.  
Escriba 'license()' o 'licence()' para detalles de distribucion.  
  
R es un proyecto colaborativo con muchos contribuyentes.  
Escriba 'contributors()' para obtener más información y  
'citation()' para saber cómo citar R o paquetes de R en publicaciones.  
  
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,  
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.  
Escriba 'q()' para salir de R.  
  
> |
```

**Recap...**

REALIDAD

Recap...

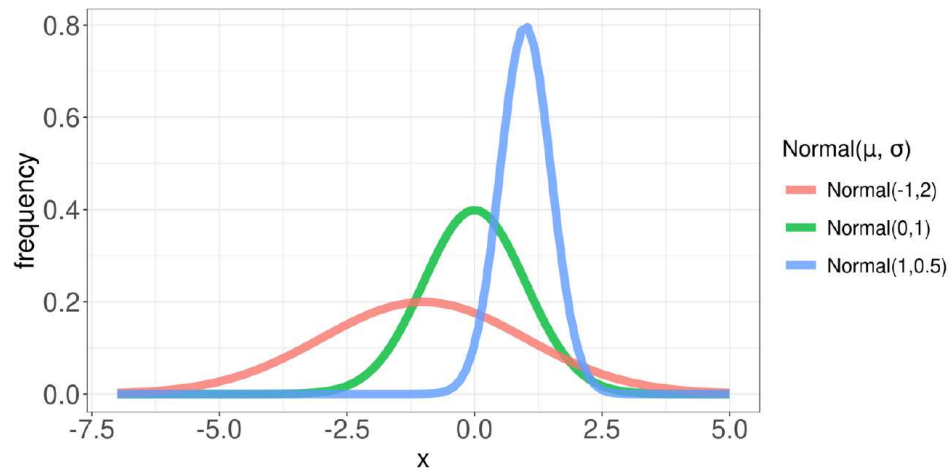
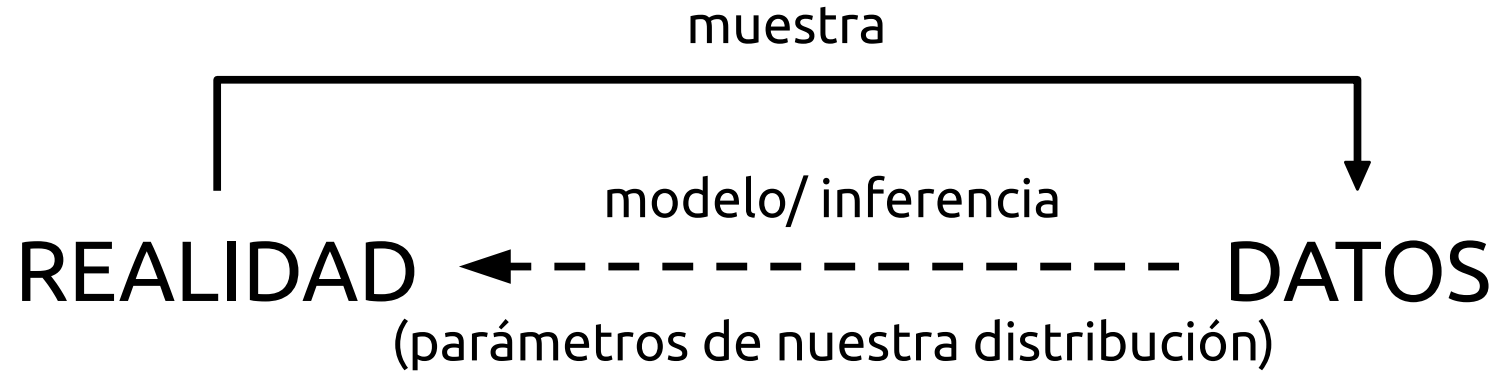


<i>i</i>	<i>peso</i>	<i>latitud</i>
1	27	40
2	35	51
3	42	59
...	...	...
n	50	65



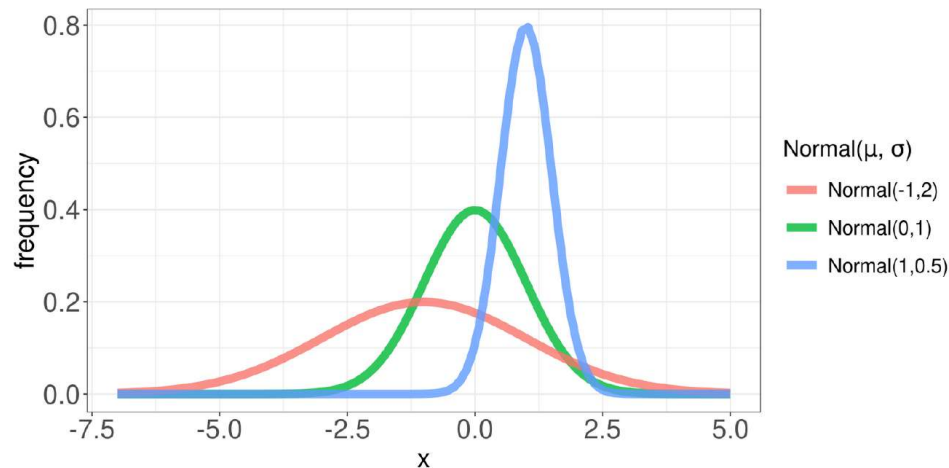
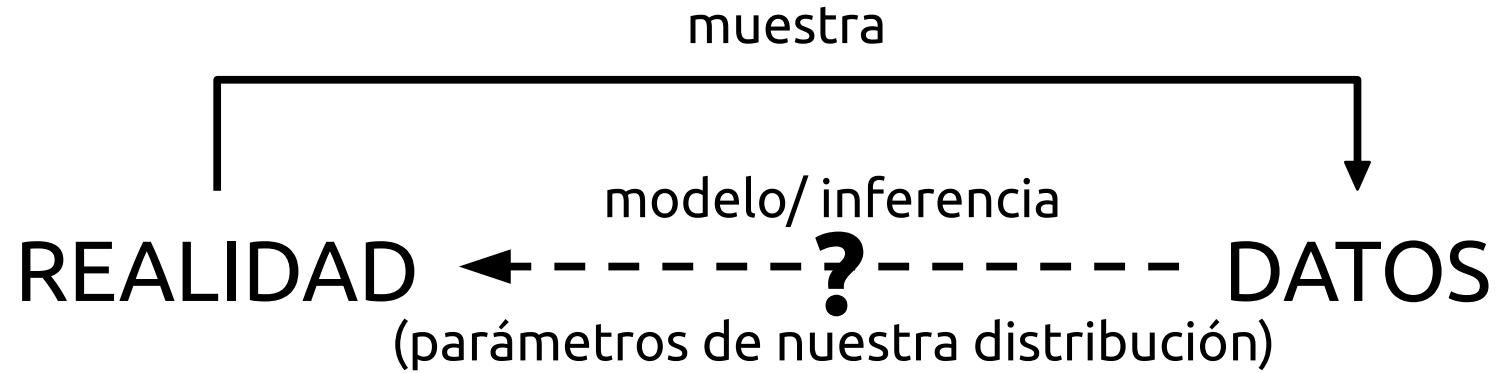
## Recap...

$i$	$peso$	$latitud$
1	27	40
2	35	51
3	42	59
...	...	...
$n$	50	65

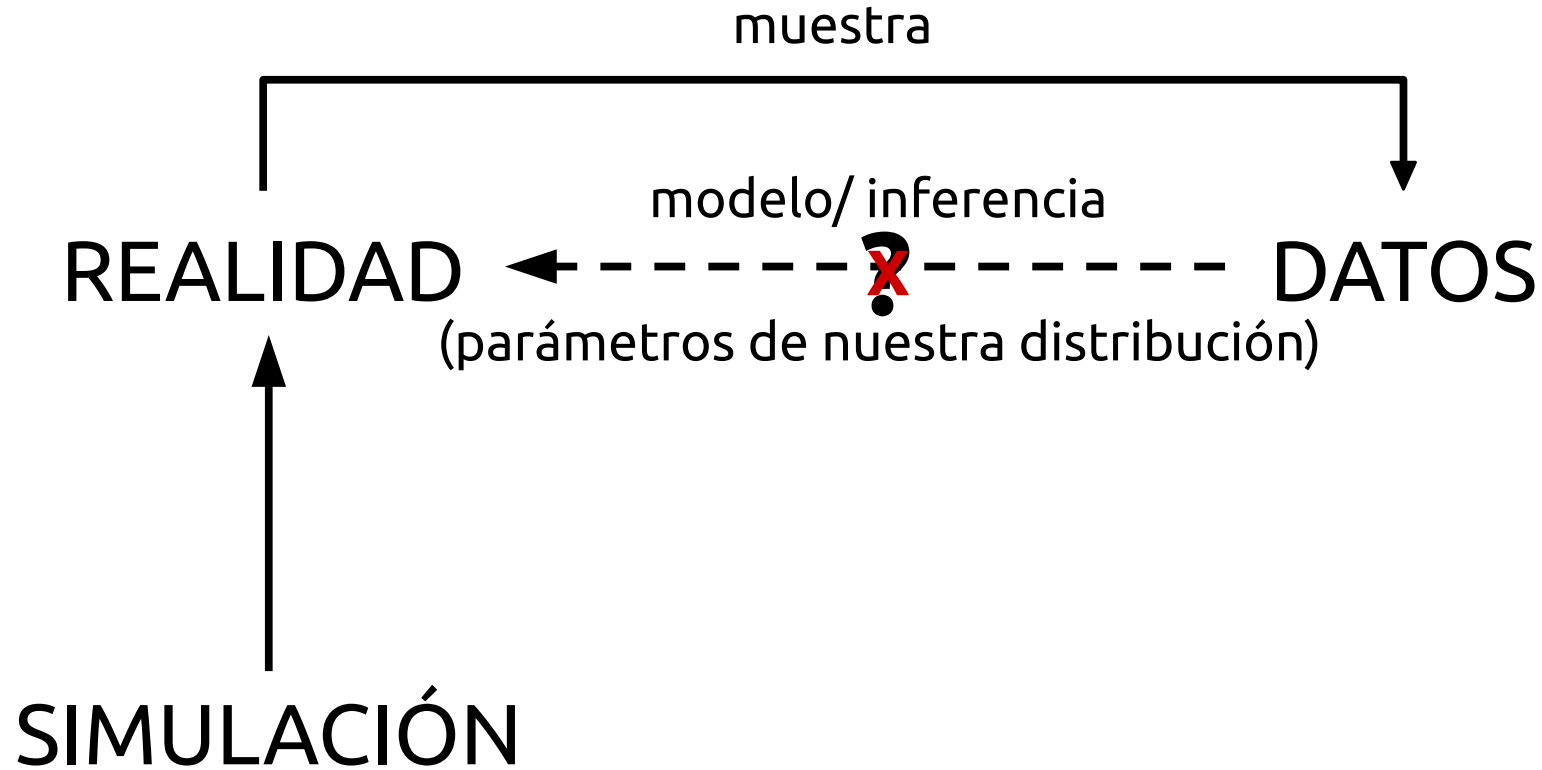


## Recap...

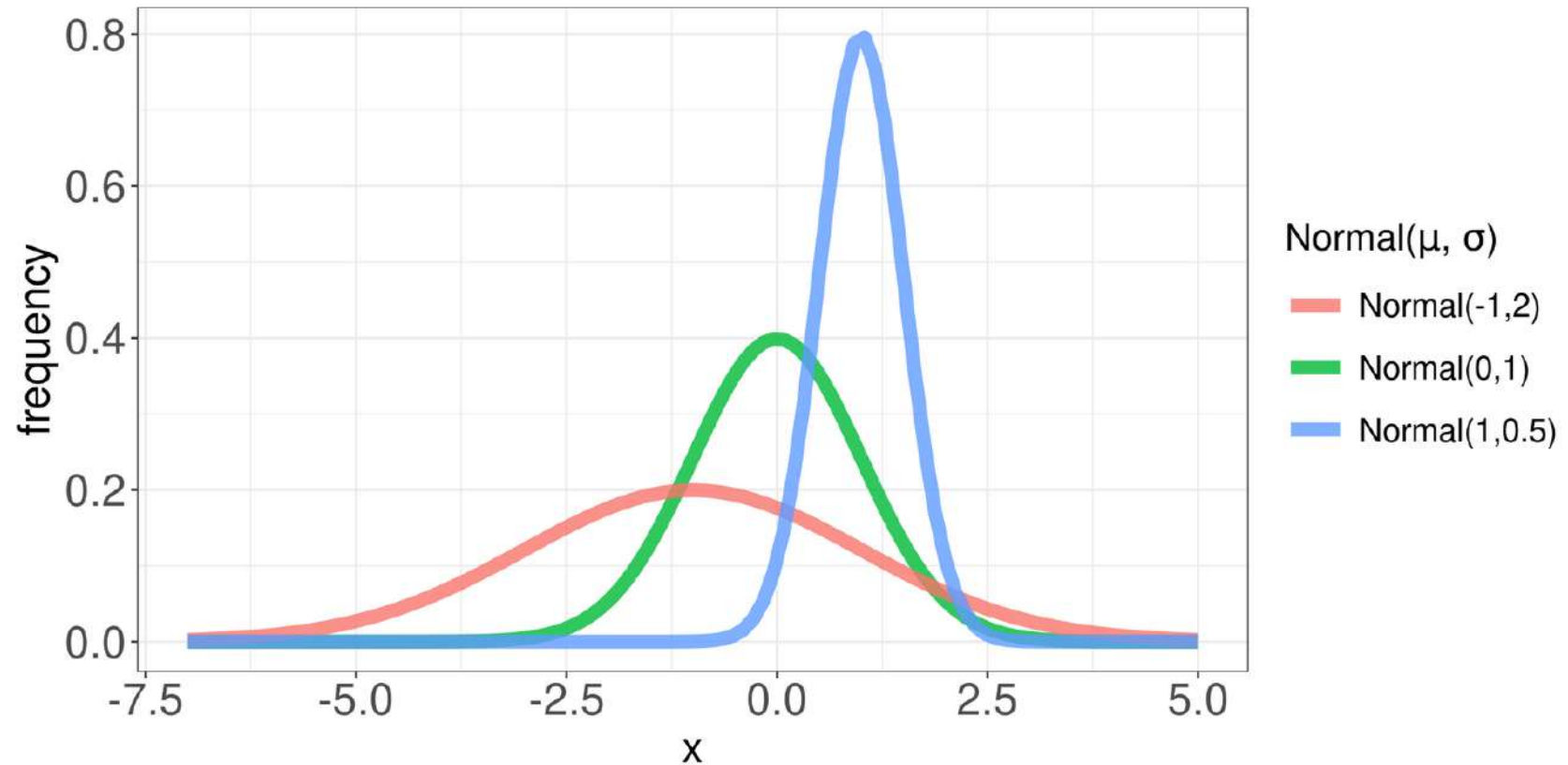
$i$	$peso$	$latitud$
1	27	40
2	35	51
3	42	59
...	...	...
$n$	50	65



## Recap...



## Nuestros modelos...



## El modelo lineal (o cuando “la media se mueve”)

$i$	$\text{peso}$
1	27
2	22
3	31
...	...
n	28

$$\text{weight}(\text{kg}) \sim \text{Normal}(\text{mean} = 27, \text{sd} = 3)$$



## El modelo lineal (o cuando “la media se mueve”)

$i$	$\text{peso}$
1	27
2	22
3	31
...	...
n	28

$$\text{weight}(\text{kg}) \sim \text{Normal}(\text{mean} = 27, \text{sd} = 3)$$

$$Y \sim \text{Normal}(\mu, \sigma)$$

$$Y \sim \text{Normal}(27, 3)$$



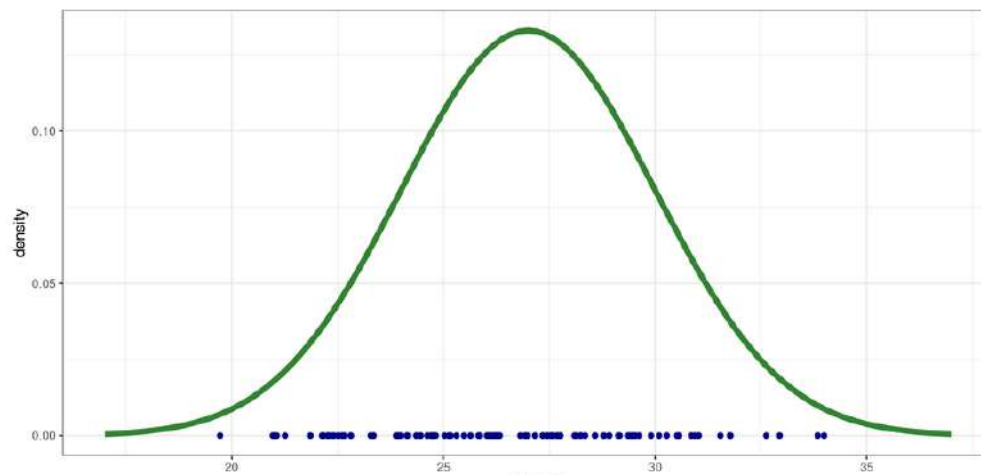
## El modelo lineal (o cuando “la media se mueve”)

$i$	$\text{peso}$
1	27
2	22
3	31
...	...
n	28

$$\text{weight}(\text{kg}) \sim \text{Normal}(\text{mean} = 27, \text{sd} = 3)$$

$$Y \sim \text{Normal}(\mu, \sigma)$$

$$Y \sim \text{Normal}(27, 3)$$



## El modelo lineal (o cuando “la media se mueve”)

$i$	$peso$	$latitud$
1	27	40
2	35	51
3	42	59
...	...	...
n	50	65

$$weight_i \sim Normal(\mu_i, \sigma)$$





## El modelo lineal (o cuando “la media se mueve”)

$i$	$peso$	$latitud$
1	27	40
2	35	51
3	42	59
...	...	...
n	50	65

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 X_i$$



## El modelo lineal (o cuando “la media se mueve”)

$i$	$\text{peso}$	$\text{latitud}$
1	27	40
2	35	51
3	42	59
...	...	...
n	50	65

$$\text{weight}_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 X_i$$

$$\begin{aligned} a + bX_i \\ \alpha + \beta X_i \end{aligned}$$



## El modelo lineal (o cuando “la media se mueve”)

$i$	$peso$	$latitud$
1	27	40
2	35	51
3	42	59
...	...	...
n	50	65

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 X_i$$

$$\mu_i = \beta_0 + \beta_1 Latitude_i$$



## El modelo lineal (o cuando “la media se mueve”)

$i$	$peso$	$latitud$
1	27	40
2	35	51
3	42	59
...	...	...
n	50	65

Variable respuesta o dependiente

$$\boxed{weight_i} \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 X_i$$

$$\mu_i = \beta_0 + \beta_1 Latitude_i$$



## El modelo lineal (o cuando “la media se mueve”)

$i$	$peso$	$latitud$
1	27	40
2	35	51
3	42	59
...	...	...
n	50	65

Variable respuesta o dependiente

$$\boxed{weight_i} \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 X_i$$

$$\mu_i = \beta_0 + \beta_1 \boxed{Latitude_i}$$

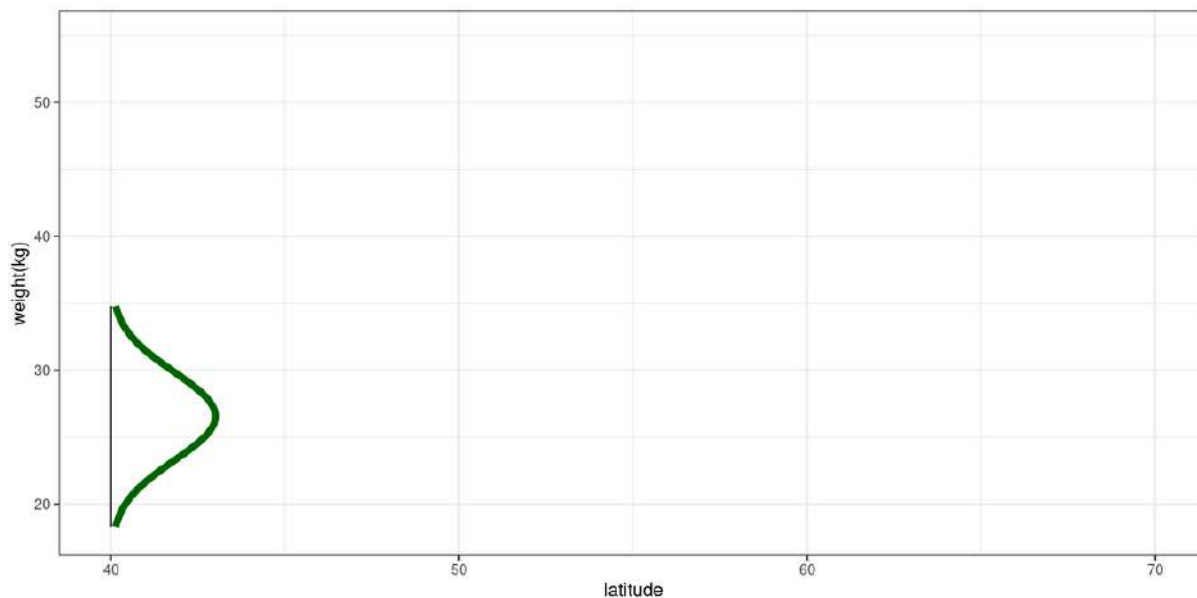
Variable o covariable  
predictora o independiente



## El modelo lineal (o cuando “la media se mueve”)

$$weight_i \sim Normal(\mu_i, \sigma)$$

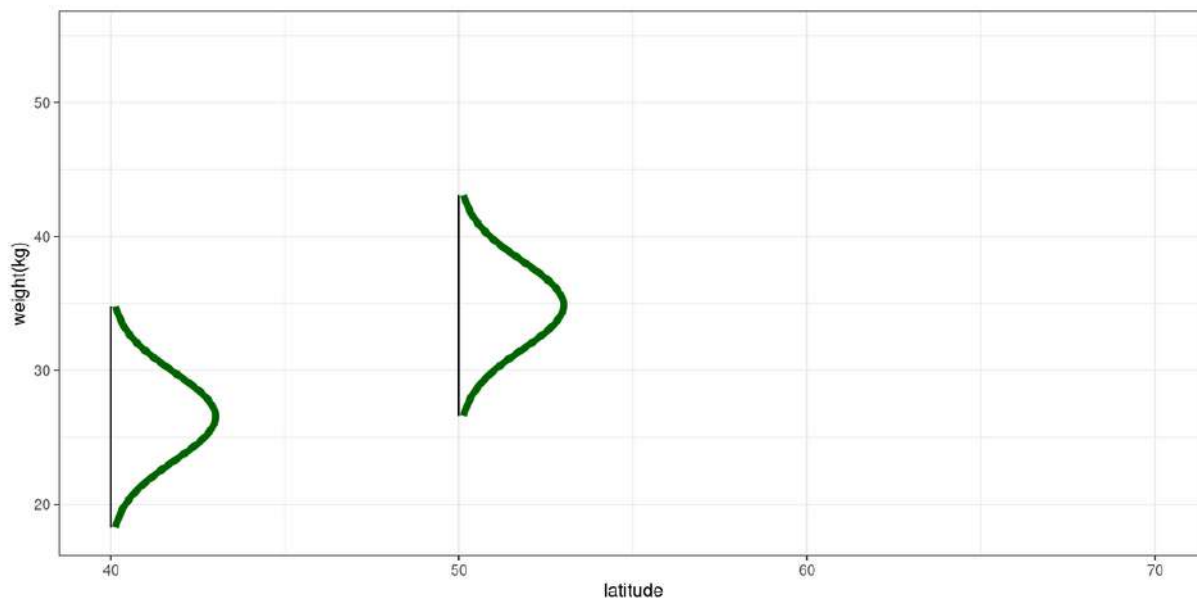
$$\mu_i = \beta_0 + \beta_1 Latitude_i$$



## El modelo lineal (o cuando “la media se mueve”)

$$weight_i \sim Normal(\mu_i, \sigma)$$

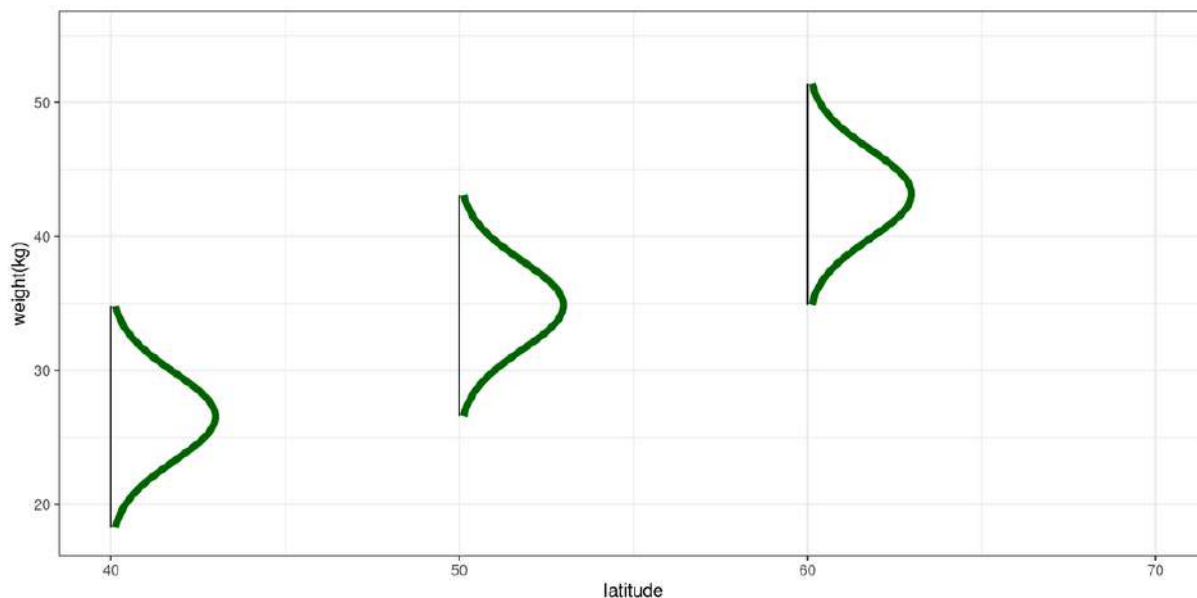
$$\mu_i = \beta_0 + \beta_1 Latitude_i$$



## El modelo lineal (o cuando “la media se mueve”)

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 Latitude_i$$

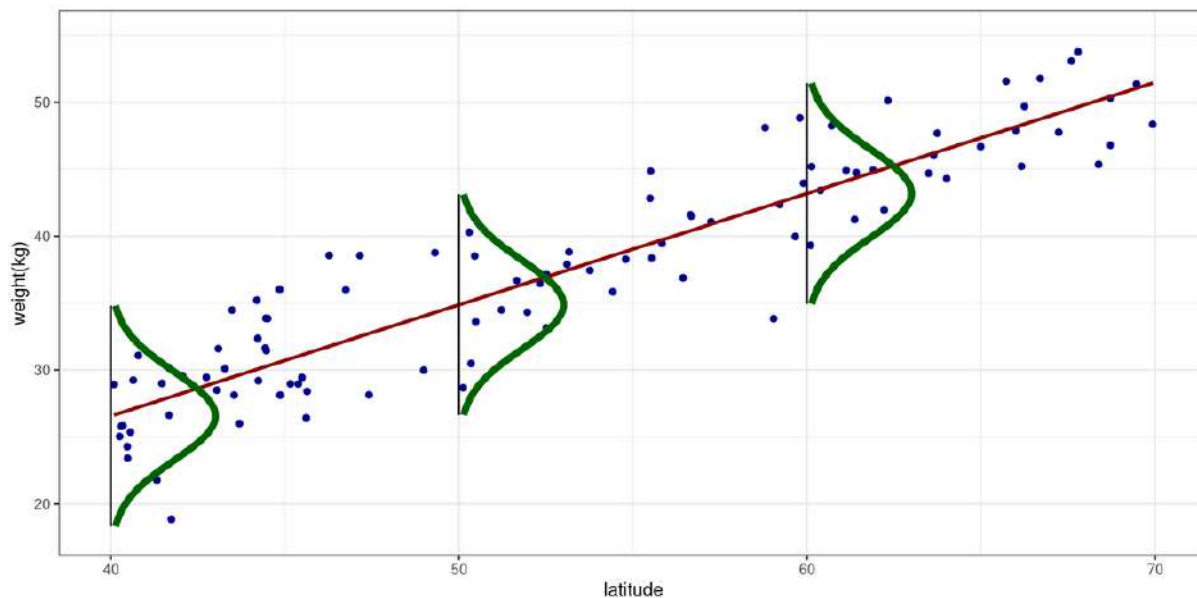




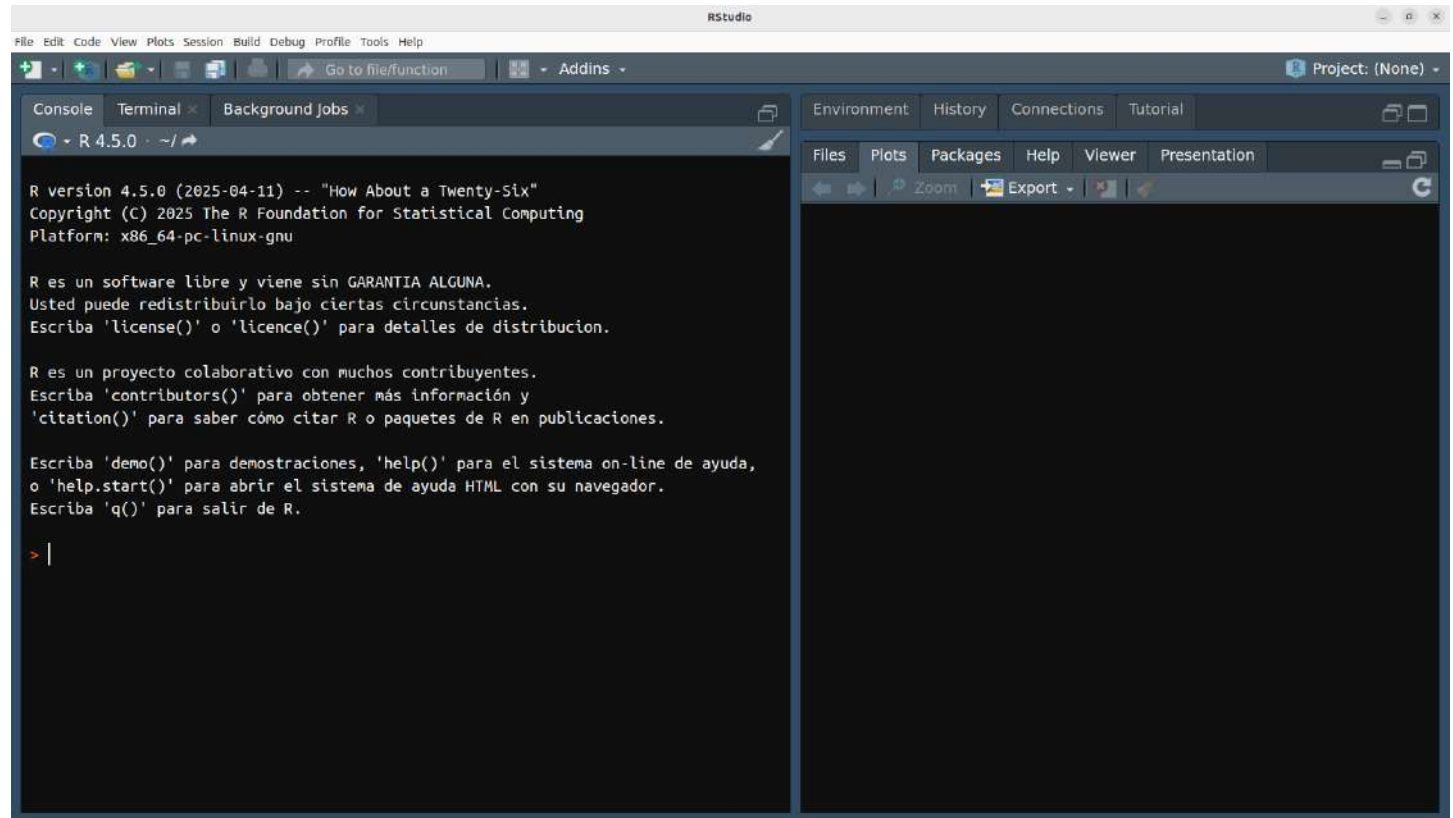
## El modelo lineal (o cuando “la media se mueve”)

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 Latitude_i$$



## El modelo lineal (o cuando “la media se mueve”)

A screenshot of the RStudio console window. The title bar says 'RStudio'. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar has icons for running code, saving, and other functions. The console shows the R startup message for version 4.5.0 (2025-04-11) on a Linux platform. The text in the console is as follows:

```
R version 4.5.0 (2025-04-11) -- "How About a Twenty-Six"  
Copyright (C) 2025 The R Foundation for Statistical Computing  
Platform: x86_64-pc-linux-gnu  
  
R es un software libre y viene sin GARANTIA ALGUNA.  
Usted puede redistribuirlo bajo ciertas circunstancias.  
Escriba 'license()' o 'licence()' para detalles de distribucion.  
  
R es un proyecto colaborativo con muchos contribuyentes.  
Escriba 'contributors()' para obtener más información y  
'citation()' para saber cómo citar R o paquetes de R en publicaciones.  
  
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,  
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.  
Escriba 'q()' para salir de R.  
  
> |
```

## El modelo lineal

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 Latitude_i$$

## El modelo lineal

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \boxed{\beta_0} + \beta_1 Latitude_i$$

- **Intercepto ( $\beta_0$ ):** Valor esperado de la variable respuesta cuando el predictor vale cero. En muchas ocasiones se interpreta como un valor de referencia matemático, no literal.

## El modelo lineal

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \boxed{\beta_1} Latitude_i$$

- **Intercepto ( $\beta_0$ ):** Valor esperado de la variable respuesta cuando el predictor vale cero. En muchas ocasiones se interpreta como un valor de referencia matemático, no literal.
- ***Estimates* o coeficientes ( $\beta_1$ ):** tasa de cambio esperada en la variable respuesta por una unidad de aumento en el predictor. Cada aumento de 1 grado de latitud se asocia con un incremento promedio de 0.8 kg en el peso esperado. Es la pendiente de la recta ( $\Delta y / \Delta x$ ).

## El modelo lineal: asunciones

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 Latitude_i$$

- No hay heterogeneidad no explicada (no faltan predictores).

## El modelo lineal: asunciones

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 Latitude_i$$

- No hay heterogeneidad no explicada (no faltan predictores).
- Conocemos la relación entre el predictor y la respuesta (y es lineal en los parámetros).

## El modelo lineal: asunciones

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 Latitude_i$$

- No hay heterogeneidad no explicada (no faltan predictores).
- Conocemos la relación entre el predictor y la respuesta (y es lineal en los parámetros).
- Conocemos la distribución de los datos: Gaussiana o Normal.

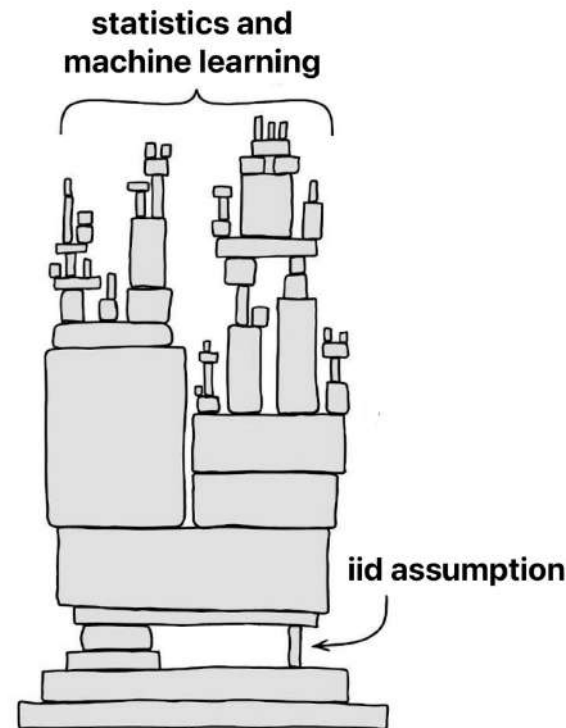


## El modelo lineal: asunciones

$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 Latitude_i$$

- No hay heterogeneidad no explicada (no faltan predictores).
- Conocemos la relación entre el predictor y la respuesta (y es lineal en los parámetros).
- Conocemos la distribución de los datos: Gaussiana o Normal.
- Los errores son *iid*: "*independent and identically distributed*", en nuestro caso con una distribución normal con media en 0.

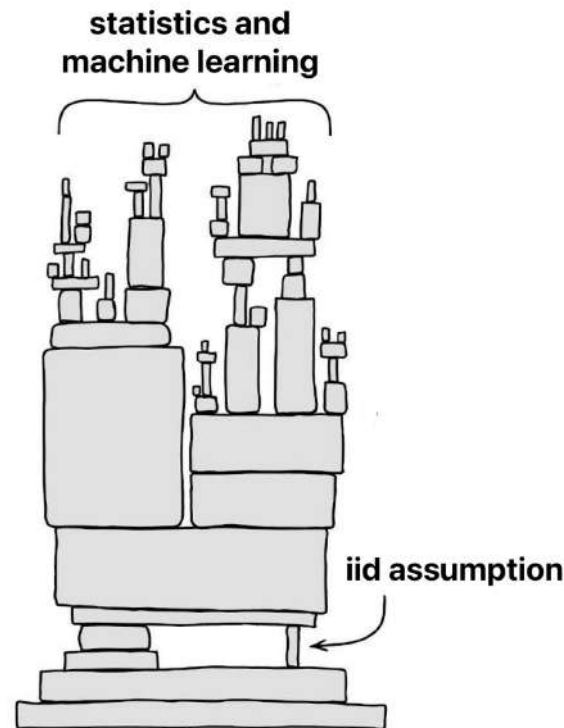


## El modelo lineal: asunciones

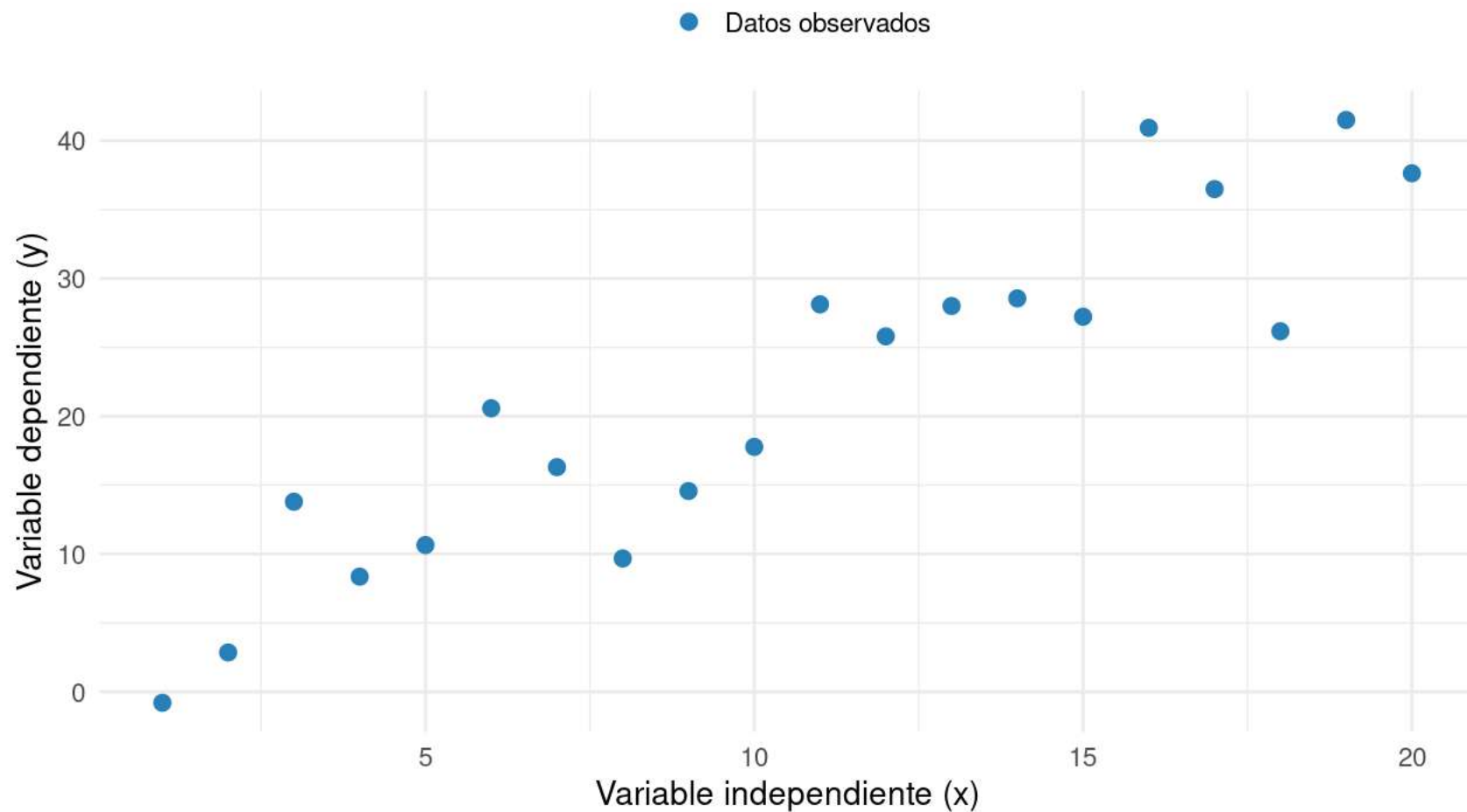
$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 Latitude_i$$

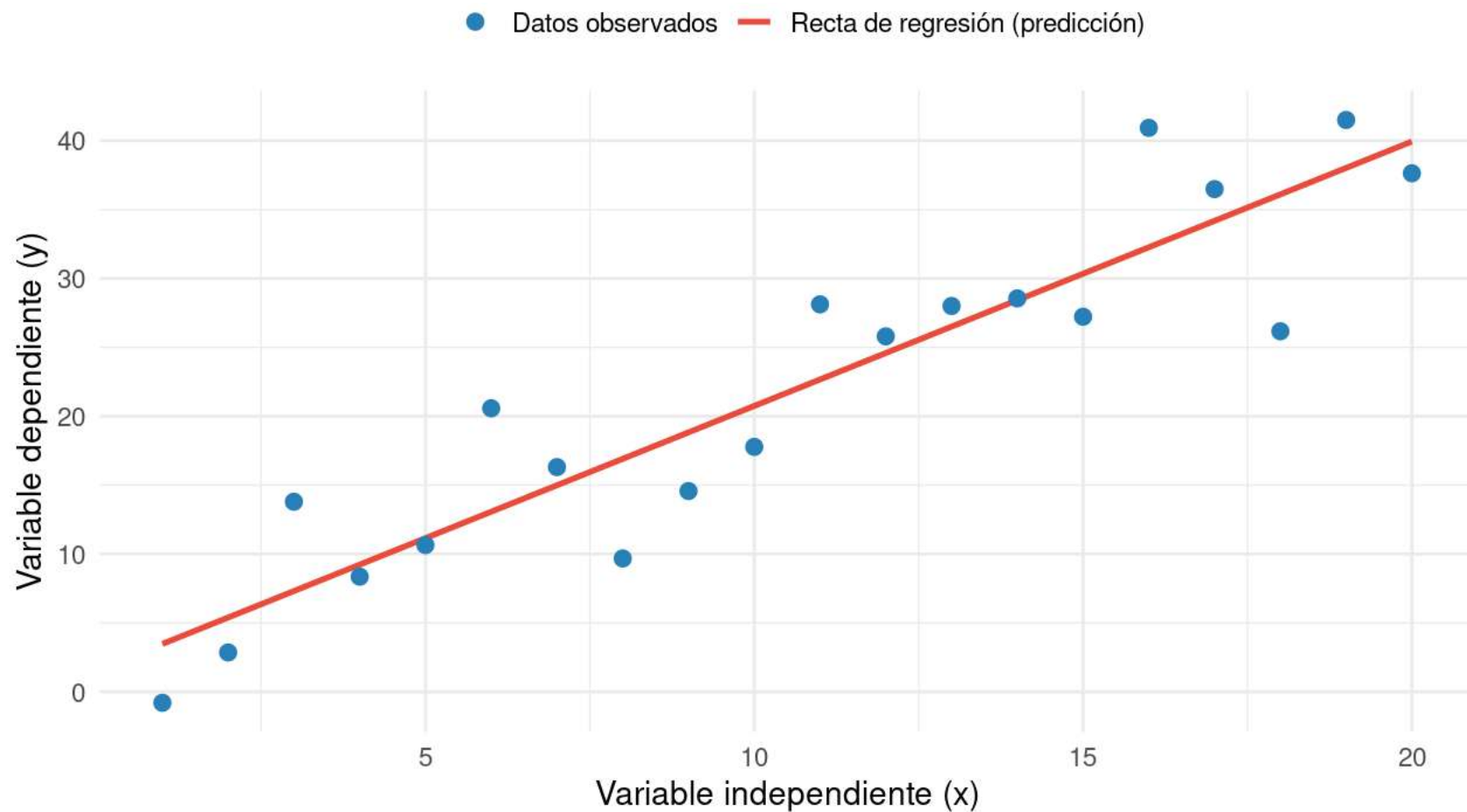
- No hay heterogeneidad no explicada (no faltan predictores).
- Conocemos la relación entre el predictor y la respuesta (y es lineal en los parámetros).
- Conocemos la distribución de los datos: Gaussiana o Normal.
- Los errores son *iid*: "*independent and identically distributed*", en nuestro caso con una distribución normal con media en 0.
- La varianza se mantiene constante: homocedasticidad.



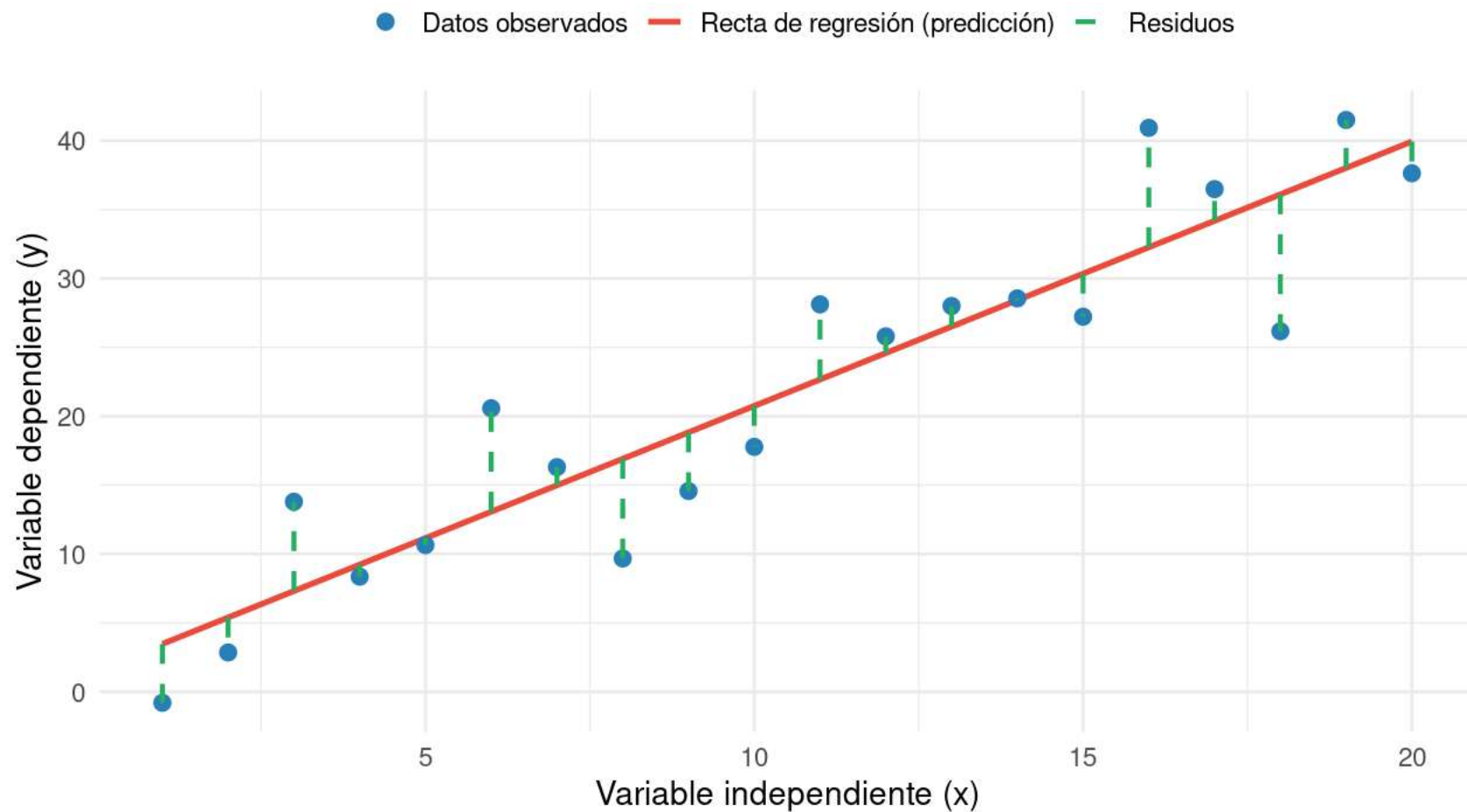
## El modelo lineal: los residuos



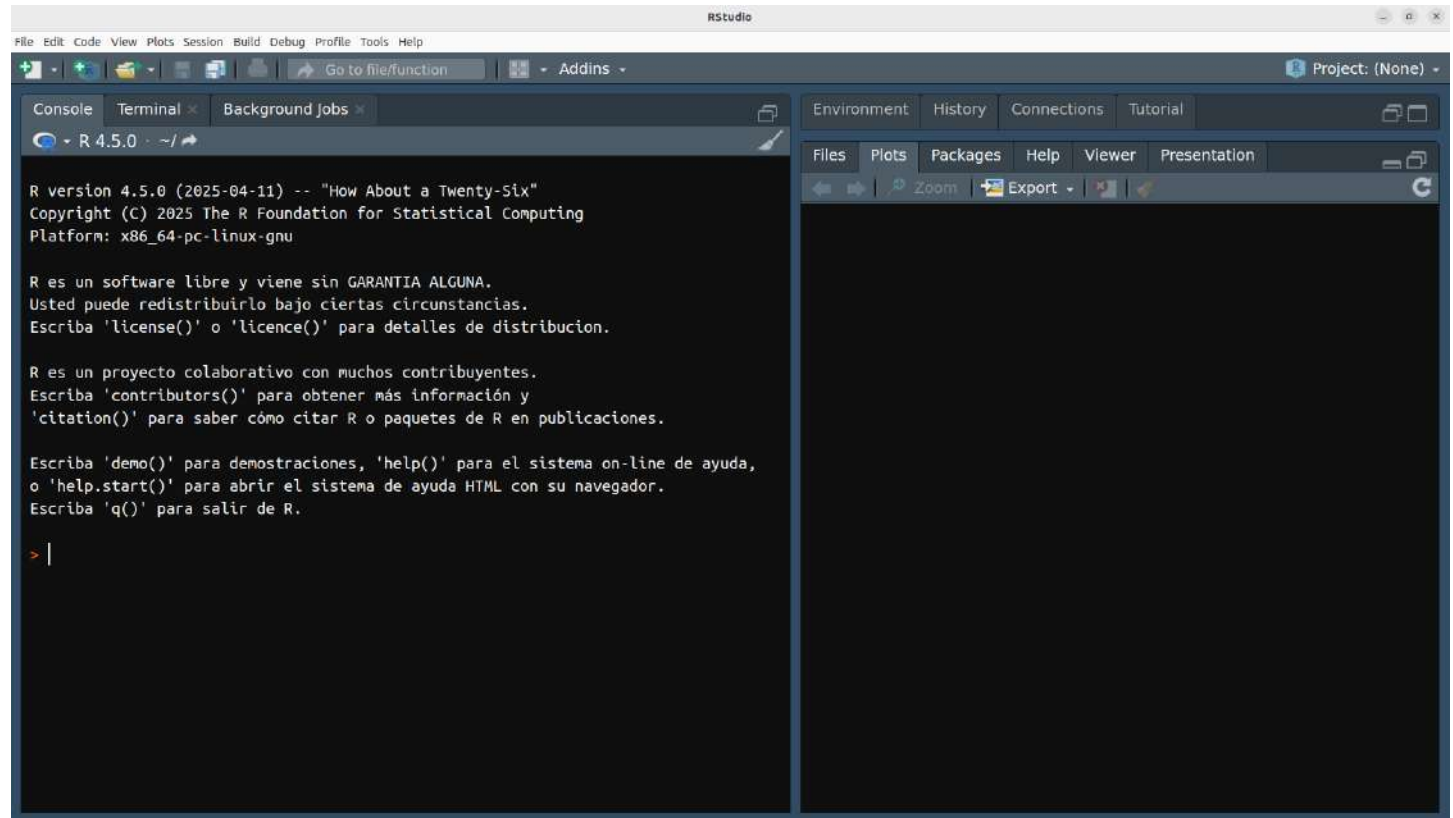
## El modelo lineal: los residuos



## El modelo lineal: los residuos



## El modelo lineal: los residuos

A screenshot of the RStudio console window. The title bar says 'RStudio'. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar has icons for running code, saving, and other functions. The console shows the R startup message for version 4.5.0 (2025-04-11) on a Linux platform. It includes copyright information and instructions on how to use R, such as running 'license()' for details, 'contributors()' for more information, and 'demo()' for demonstrations. The prompt '>' is visible at the bottom of the console.

```
R version 4.5.0 (2025-04-11) -- "How About a Twenty-Six"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

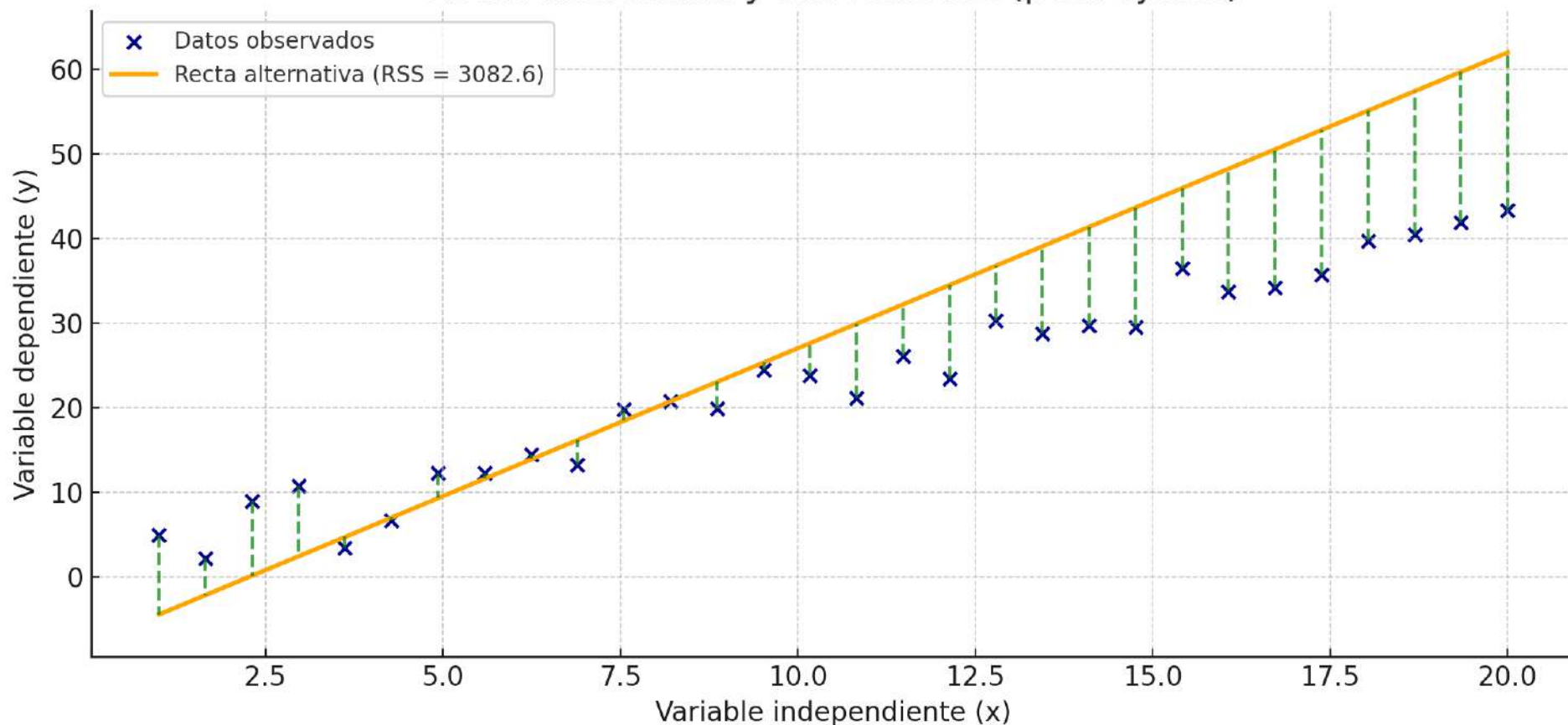
R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> |
```

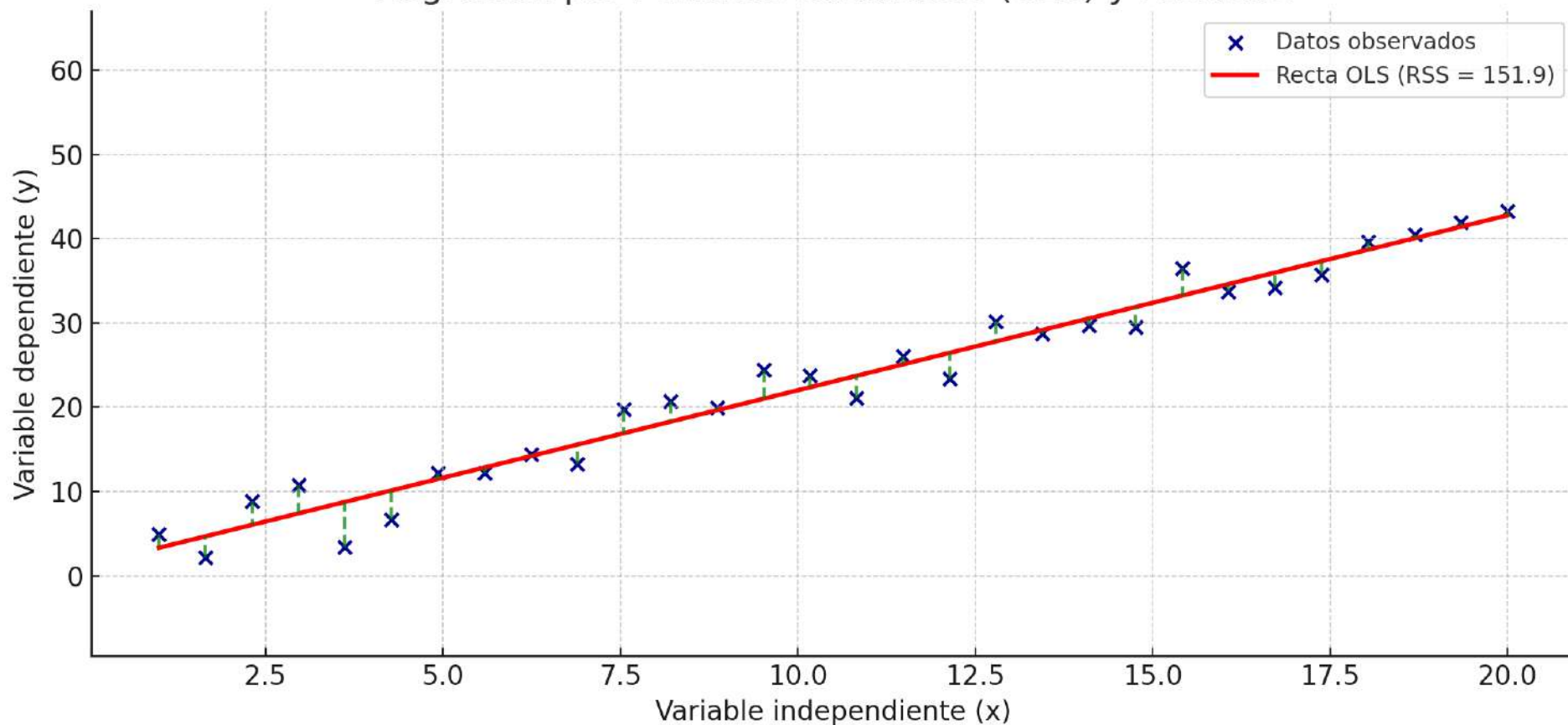
## El modelo lineal: mínimos cuadrados ordinarios (OLS)

Recta alternativa y sus residuos (peor ajuste)



## El modelo lineal: mínimos cuadrados ordinarios (OLS)

Regresión por Mínimos Cuadrados (OLS) y residuos





### El modelo lineal: mínimos cuadrados ordinarios (OLS)

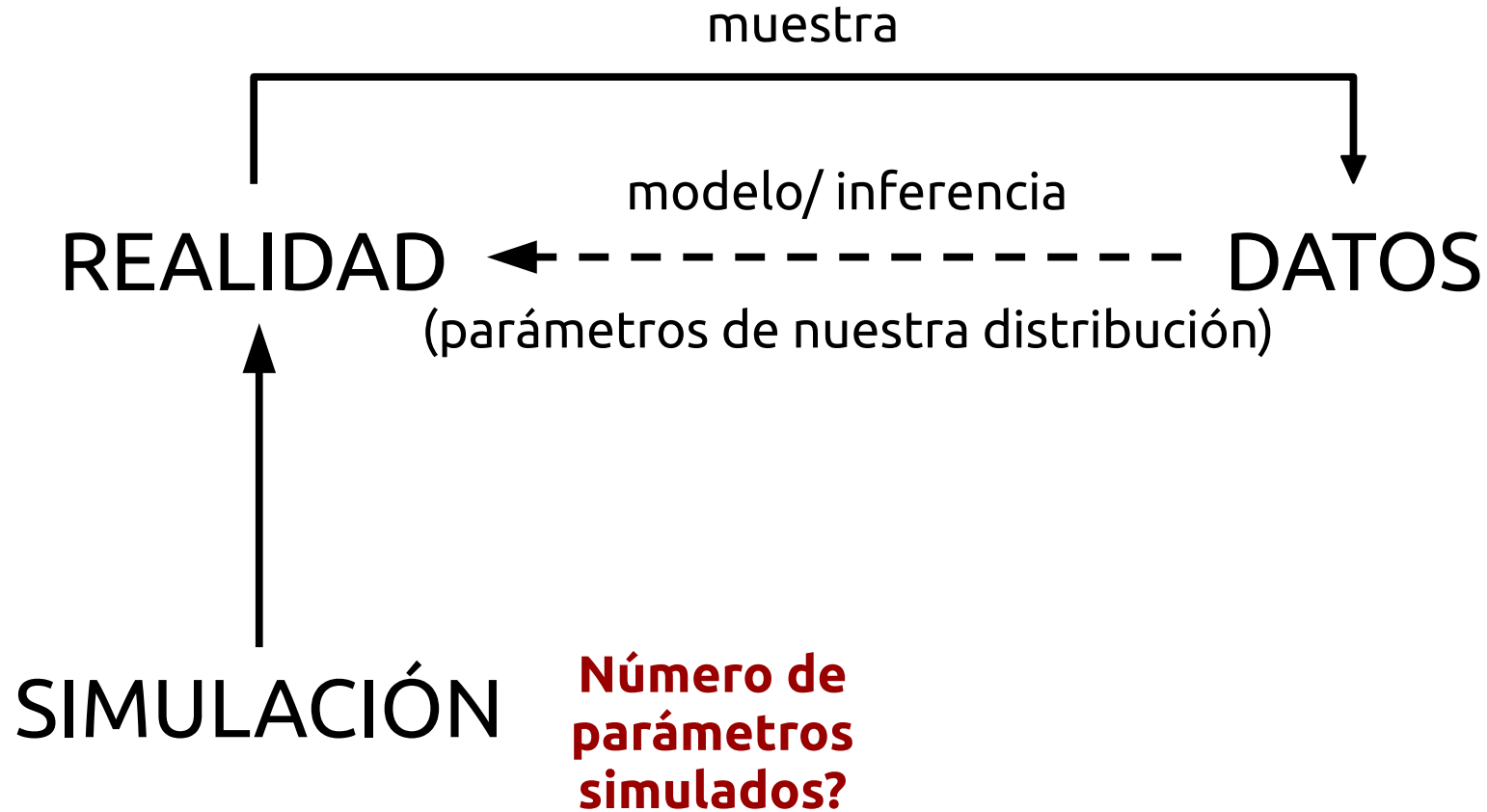
- El método de mínimos cuadrados ordinarios (*ordinary least squares*, OLS) busca estimar los parámetros  $\beta_0, \beta_1, \dots, \beta_n$  de un modelo lineal de forma que los valores ajustados o predichos estén lo más cerca posible de los valores observados.

## El modelo lineal: mínimos cuadrados ordinarios (OLS)

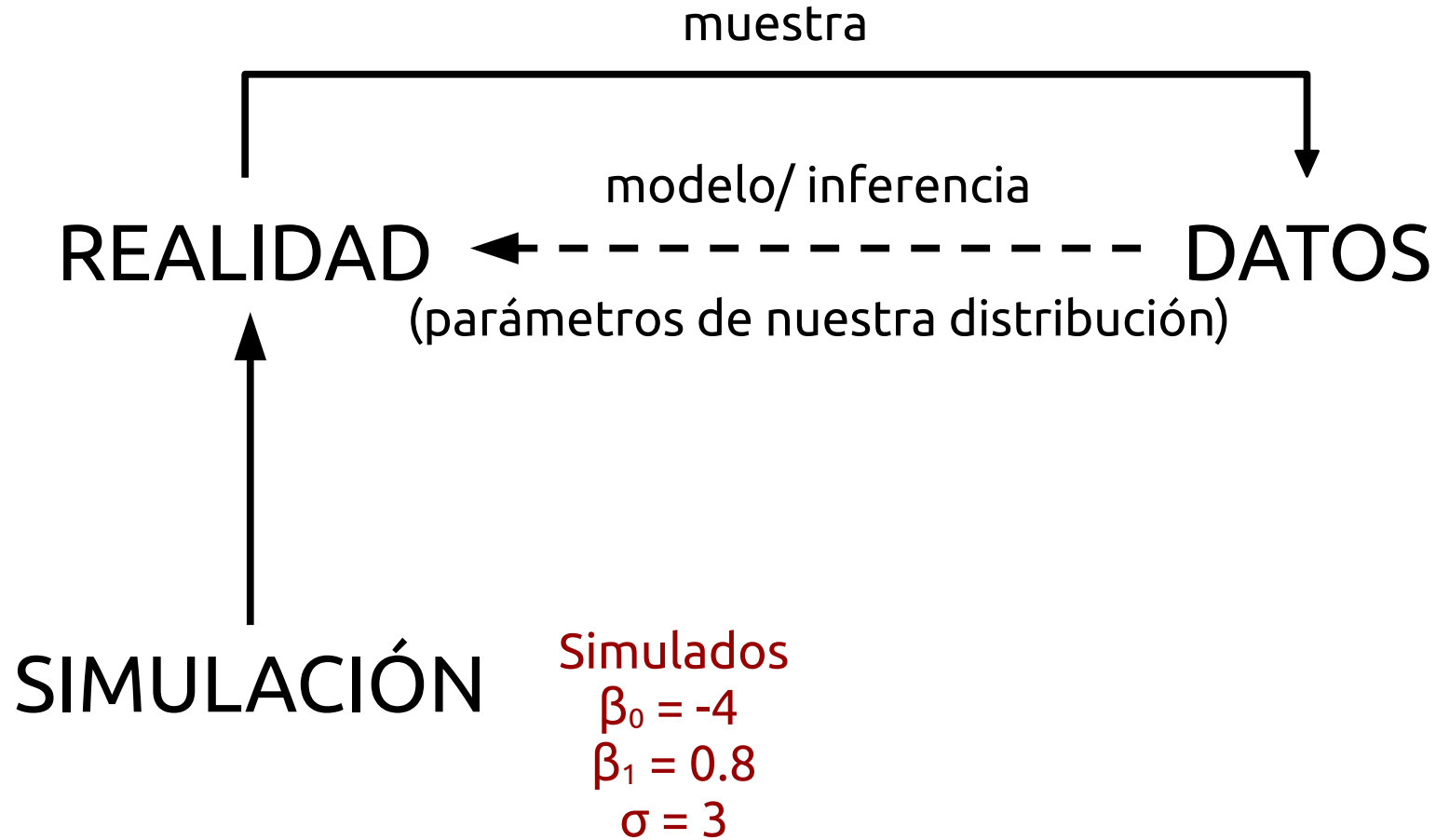
- El método de mínimos cuadrados ordinarios (*ordinary least squares*, OLS) busca estimar los parámetros  $\beta_0, \beta_1, \dots, \beta_n$  de un modelo lineal de forma que los valores ajustados o predichos estén lo más cerca posible de los valores observados.
- En otras palabras, OLS minimiza la suma de los cuadrados de los residuos:

$$\text{RSS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}))^2$$

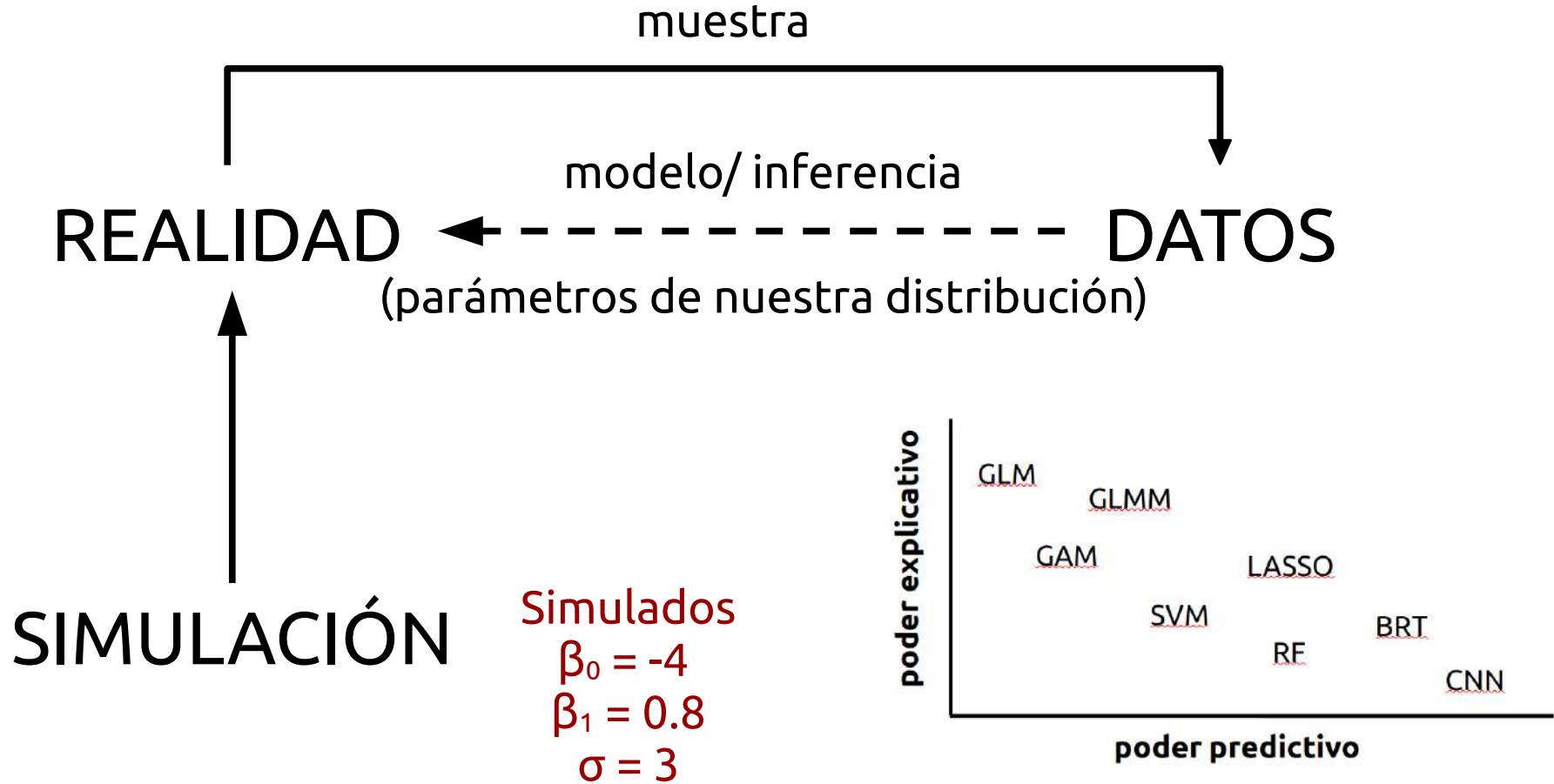
## Recap...



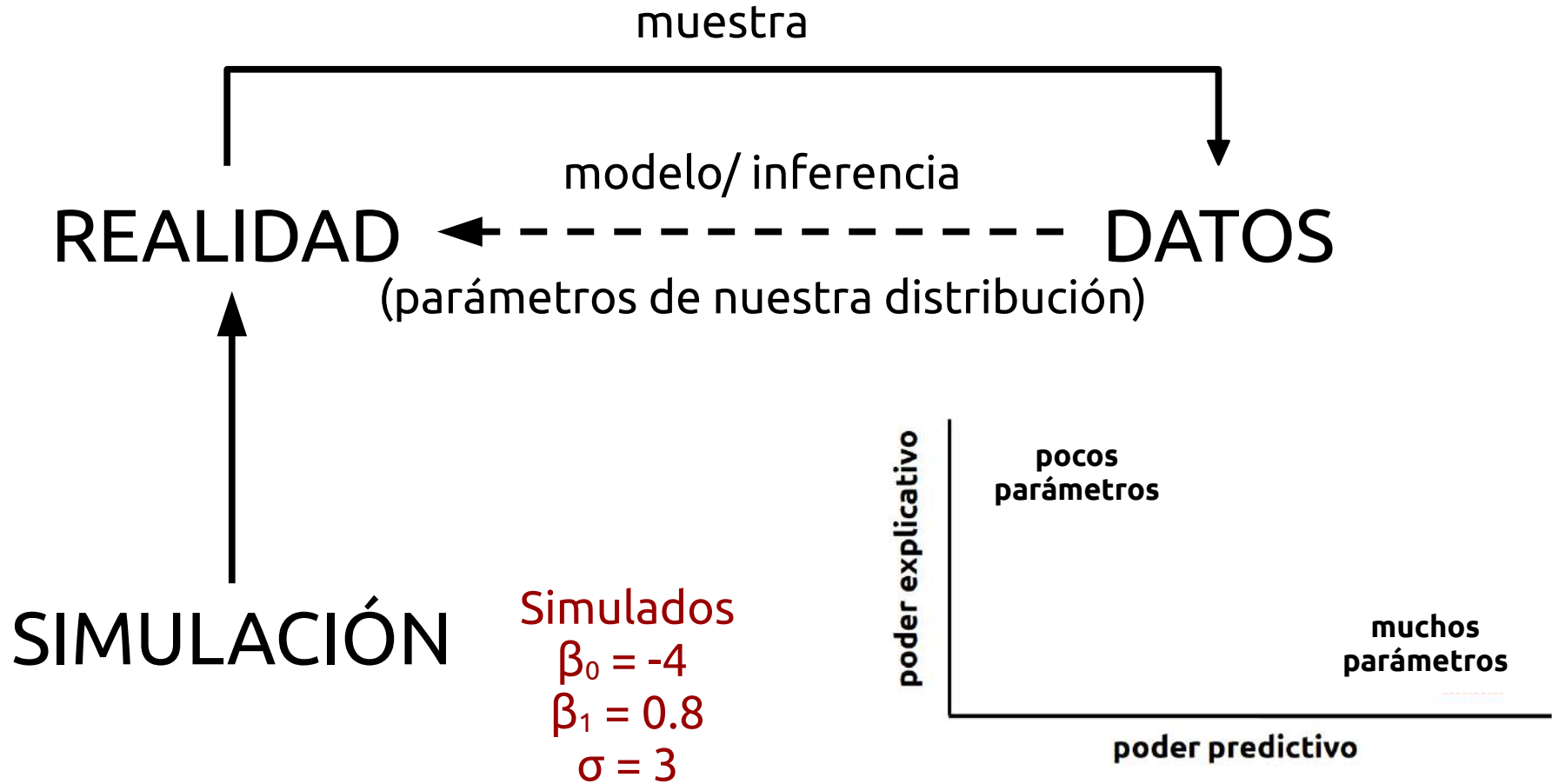
## Recap...



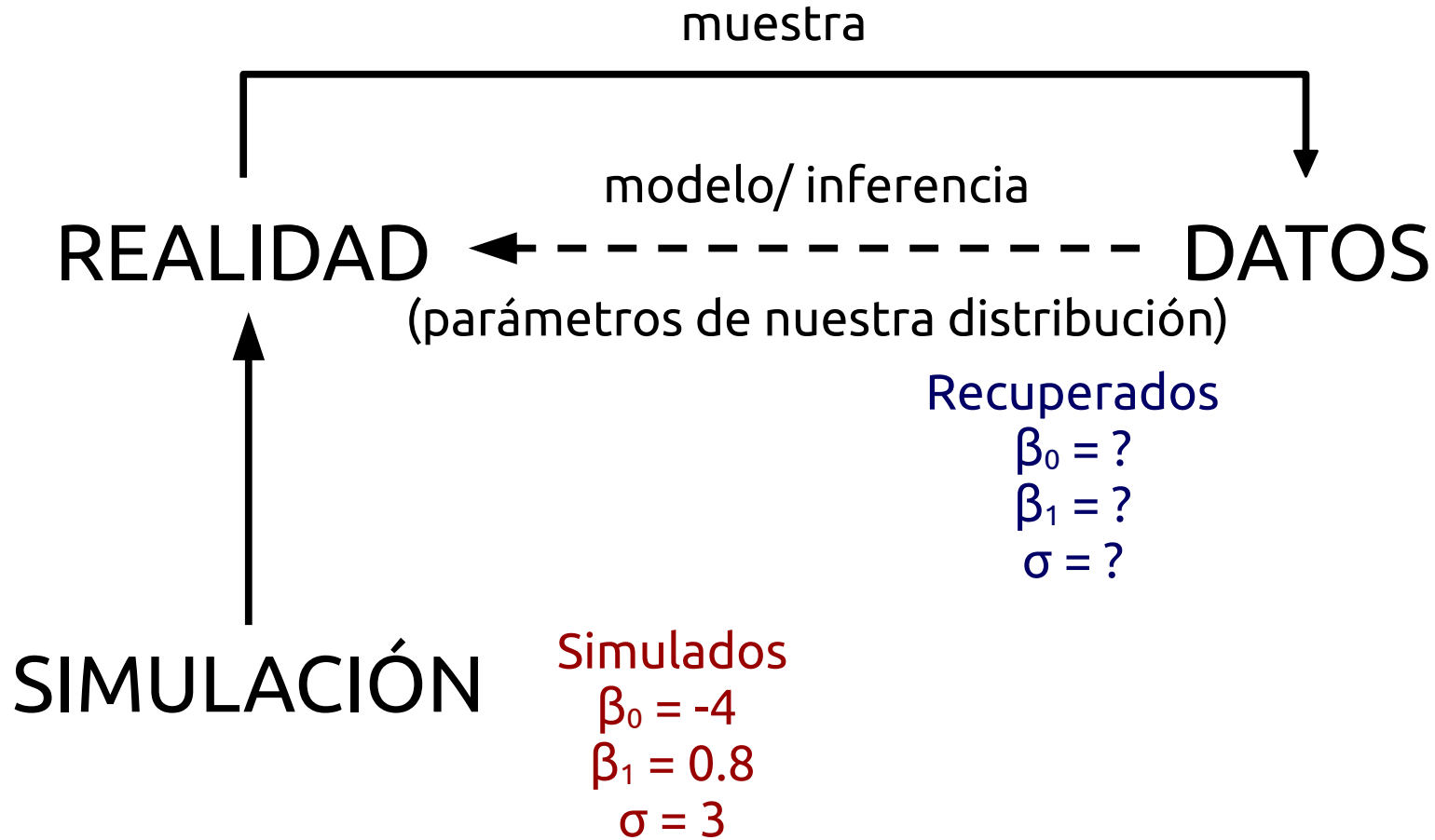
## Recap...



## Recap...



## Recap...



## Ejercicio propuesto (5 min)

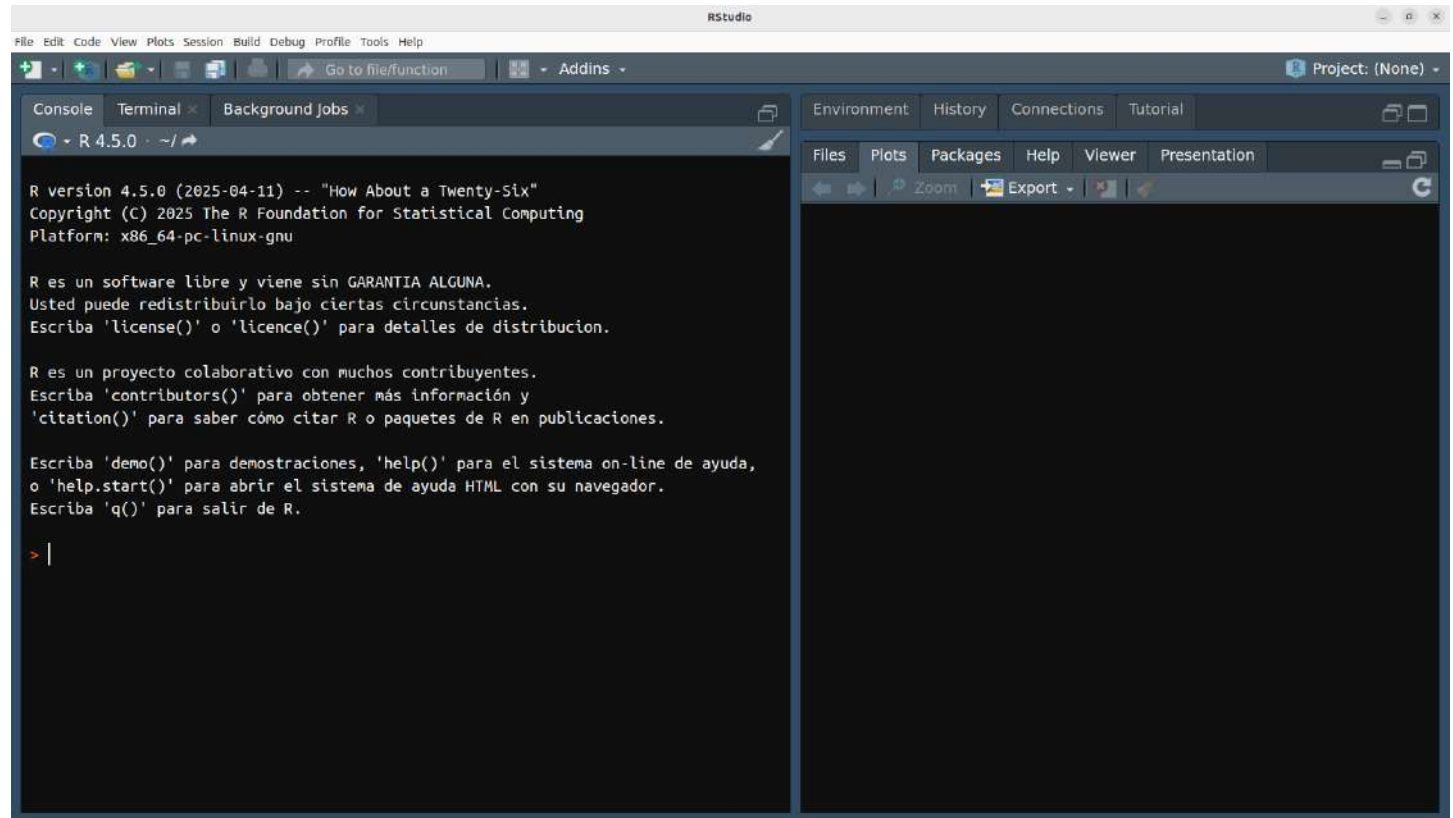
- Simula una variable predictora usando una distribución uniforme o una distribución normal, pero que tenga unos valores coherentes con lo que quieres simular (temperatura, precipitación, cobertura de bosque, peso, etc.).
- Crea un predictor lineal.
- Genera observaciones a partir de una distribución lineal cuya media sea el predictor lineal y la desviación estándar que elijas.
- Ajusta un modelo lineal usando la función `glm()`.



## Sobre el P-valor...

- El p-valor es la probabilidad de obtener unos datos tan extremos como los observados, si la hipótesis nula (el no-efecto del predictor) fuera cierta.
- “Si el predictor no tuviese efecto ( $\beta_1 = 0$  realmente), observar un resultado igual o más extremo que este ocurriría el p-valor % de las veces”

## El modelo lineal: gráficos diagnóstico

A screenshot of the RStudio application window. The title bar says 'RStudio'. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu bar is a toolbar with icons for file operations and a 'Go to file/function' search bar. The main window is divided into four panes. The top-left pane is the 'Console', showing the R startup message: 'R version 4.5.0 (2025-04-11) -- "How About a Twenty-Six"', 'Copyright (C) 2025 The R Foundation for Statistical Computing', 'Platform: x86\_64-pc-linux-gnu', and instructions on how to use R. The top-right pane is the 'Environment' pane, showing 'Project: (None)'. The bottom-left pane is the 'Files' pane, showing a list of files. The bottom-right pane is the 'Plots' pane, showing a list of plots. The console text is as follows:

```
R version 4.5.0 (2025-04-11) -- "How About a Twenty-Six"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu

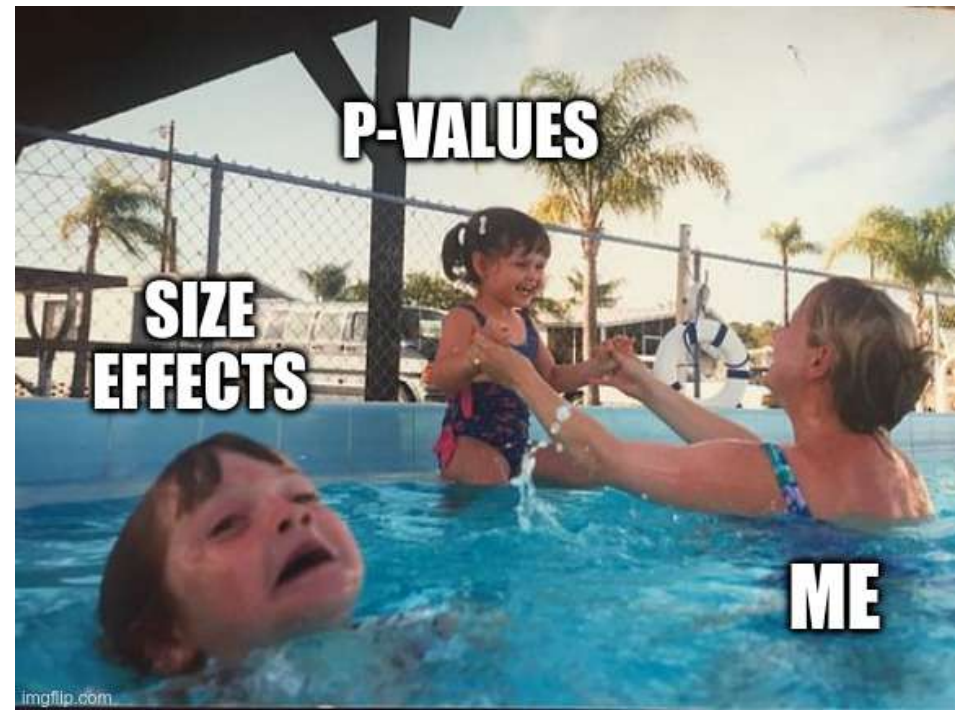
R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> |
```

## Sobre tamaño del efecto...



## Sobre tamaño del efecto...

- Es el cambio esperado en la variable respuesta por cada unidad de cambio en la variable predictora.
- En un modelo lineal simple, es literalmente la definición de “pendiente de una recta”.

## Sobre tamaño del efecto...

- Es el cambio esperado en la variable respuesta por cada unidad de cambio en la variable predictora.
- En un modelo lineal simple, es literalmente la definición de “pendiente de una recta”.

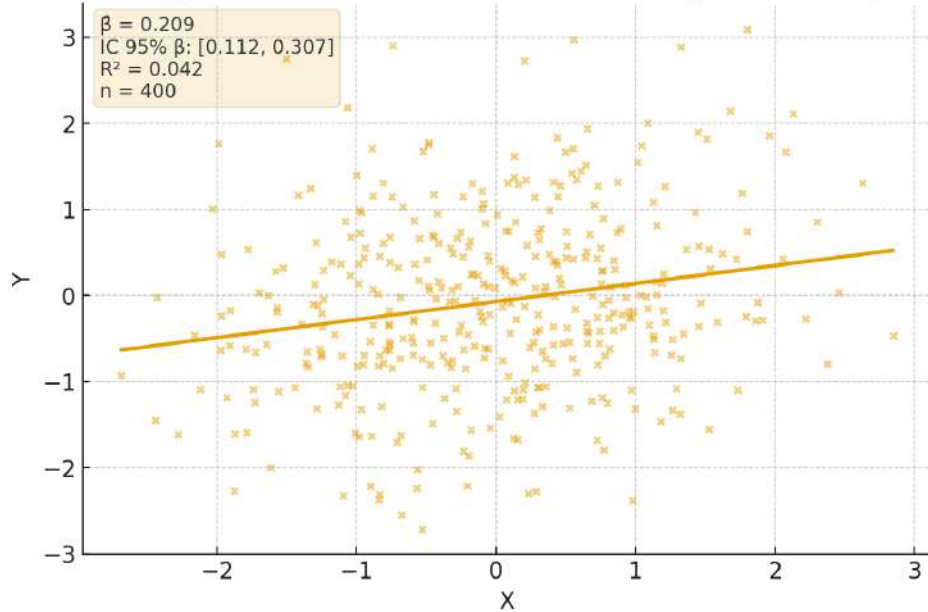
$$weight_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \boxed{\beta_1} X_i$$

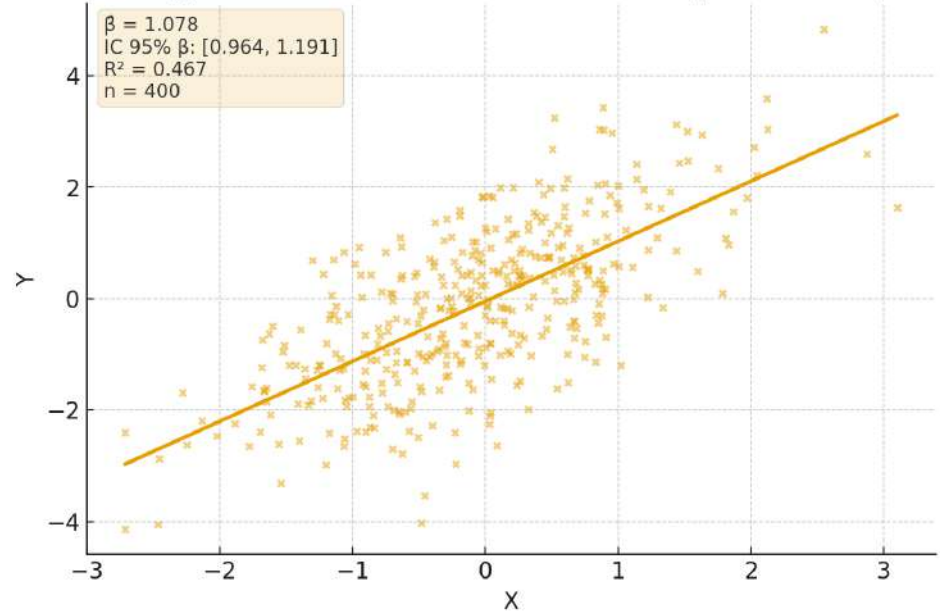
$$\boxed{\beta_1} = \frac{\Delta \mu}{\Delta X} = \frac{\Delta weight}{\Delta latitude}$$

## Sobre tamaño del efecto...

Regresión lineal: tamaño del efecto ( $\beta$  real = 0.2)



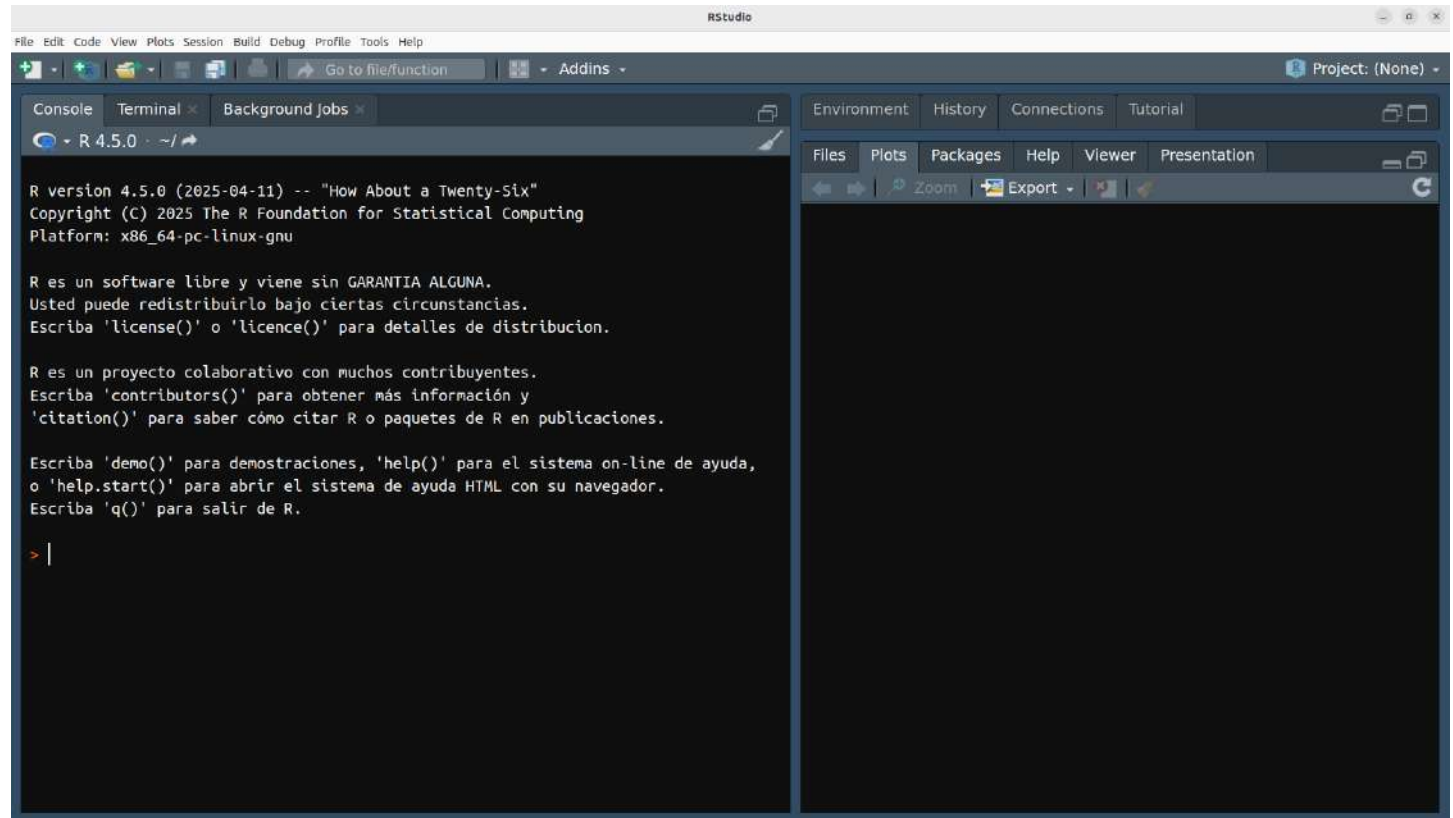
Regresión lineal: tamaño del efecto ( $\beta$  real = 1.0)



## Sobre tamaño del efecto...

- Es el cambio esperado en la variable respuesta por cada unidad de cambio en la variable predictora.
- En un modelo lineal simple, es literalmente la definición de “pendiente de una recta”.
- Es muy importante tener en cuenta las unidades en las que están los predictores, puesto que afecta directamente al coeficiente que lo acompaña.

## El modelo lineal: gráficos diagnóstico

A screenshot of the RStudio console window. The title bar says 'RStudio'. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar has icons for running code, saving, and other functions. The console shows the R startup message for version 4.5.0 (2025-04-11) on a Linux platform. The text in the console is as follows:

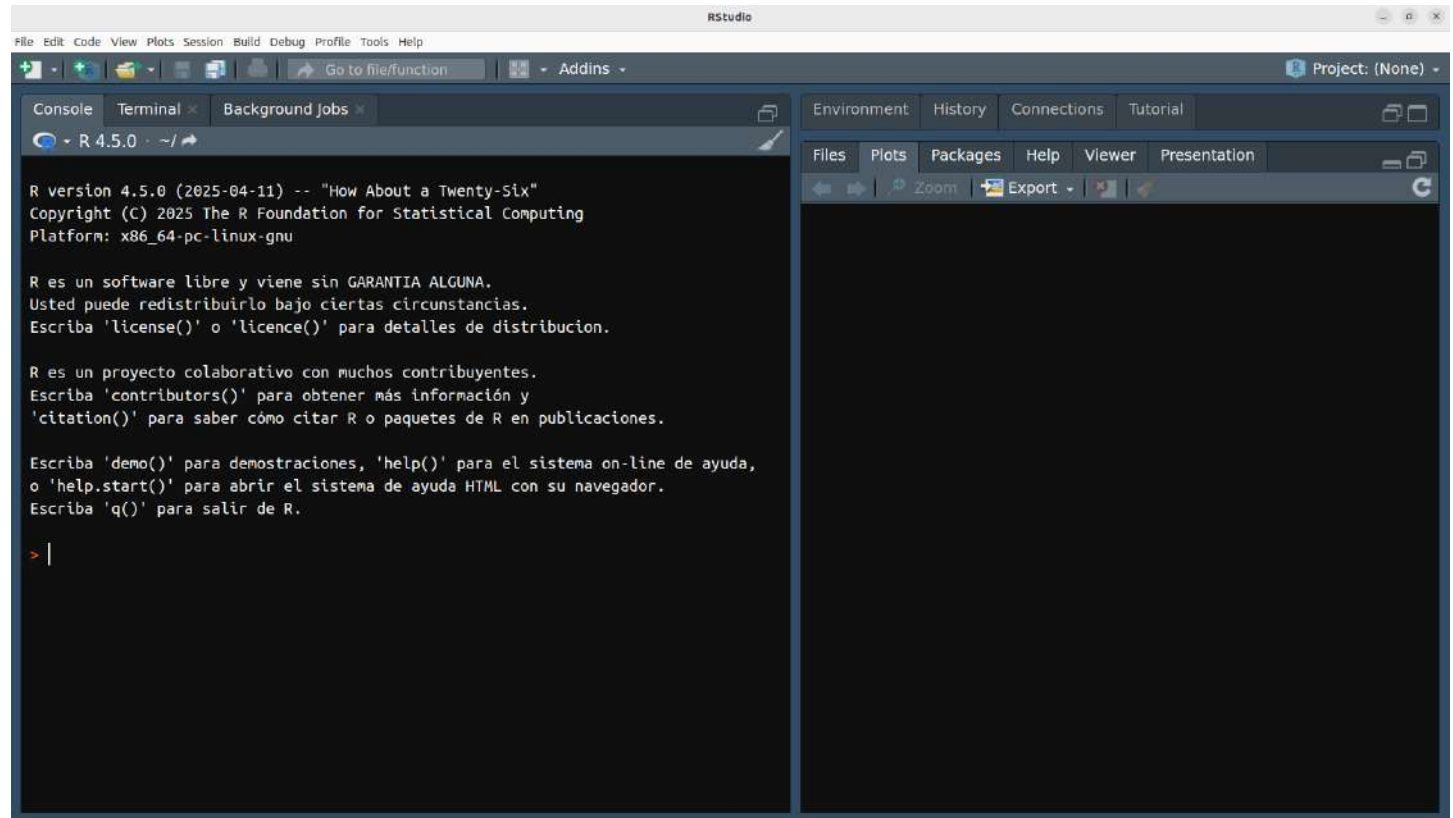
```
R version 4.5.0 (2025-04-11) -- "How About a Twenty-Six"  
Copyright (C) 2025 The R Foundation for Statistical Computing  
Platform: x86_64-pc-linux-gnu  
  
R es un software libre y viene sin GARANTIA ALGUNA.  
Usted puede redistribuirlo bajo ciertas circunstancias.  
Escriba 'license()' o 'licence()' para detalles de distribucion.  
  
R es un proyecto colaborativo con muchos contribuyentes.  
Escriba 'contributors()' para obtener más información y  
'citation()' para saber cómo citar R o paquetes de R en publicaciones.  
  
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,  
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.  
Escriba 'q()' para salir de R.  
  
> |
```



## Gráficos diagnóstico

Gráfico	Eje X	Eje Y	Qué diagnostica	Qué asunción comprueba
1 Residuals vs Fitted	Valores ajustados	Residuos	No linealidad / heterocedasticidad	Linealidad, homocedasticidad
2 Q-Q plot	Cuantiles teóricos	Cuantiles de residuos	Normalidad	Normalidad de los errores
3 Scale-Location	Valores ajustados	√	Residuos est.	
4 Residuals vs Leverage	Leverage	Residuos est.	Influencia / outliers	Independencia, influencia

## El modelo lineal: gráficos diagnóstico

A screenshot of the RStudio application window. The title bar says 'RStudio'. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu bar is a toolbar with icons for file operations and a 'Go to file/function' search bar. The main window is divided into four panes. The top-left pane is the 'Console', showing the R startup message: 'R version 4.5.0 (2025-04-11) -- "How About a Twenty-Six"', 'Copyright (C) 2025 The R Foundation for Statistical Computing', 'Platform: x86\_64-pc-linux-gnu', followed by a disclaimer in Spanish, information about contributors, and instructions on how to use help and exit. The bottom-left pane is the 'Terminal', which is currently empty. The top-right pane is the 'Environment' pane, showing 'Project: (None)'. The bottom-right pane is the 'Plots' pane, which is currently empty. The right side of the interface has a sidebar with tabs for Files, Plots, Packages, Help, Viewer, and Presentation.

## Bondad de ajuste

- Medida que indica cuánto coinciden los valores **predichos** por un modelo con los valores **observados**, reflejando su capacidad para **describir adecuadamente los datos**.

## Bondad de ajuste

- Medida que indica cuánto coinciden los valores **predichos** por un modelo con los valores **observados**, reflejando su capacidad para **describir adecuadamente los datos**.
- En un modelo lineal, podemos usar el  $R^2$  o el  $\sigma$  **residual**, que evalúa la **varianza explicada por el modelo**.

## Bondad de ajuste

- Medida que indica cuánto coinciden los valores **predichos** por un modelo con los valores **observados**, reflejando su capacidad para **describir adecuadamente los datos**.
- En un modelo lineal, podemos usar el  $R^2$  o el  $\sigma$  **residual**, que evalúa la **varianza explicada por el modelo**.
- En un modelo lineal generalizado es algo más complejo. La métrica fundamental es la **devianza**, que compara tu modelo con un modelo que fuese “perfecto”, esto es, que ajustase perfectamente todos tus datos (modelo saturado).
  - **Devianza pequeña** → tu modelo se acerca al saturado → **mejor ajuste**
  - **Devianza grande** → tu modelo se aleja del saturado → **peor reflejo de los datos**

## Bondad de ajuste

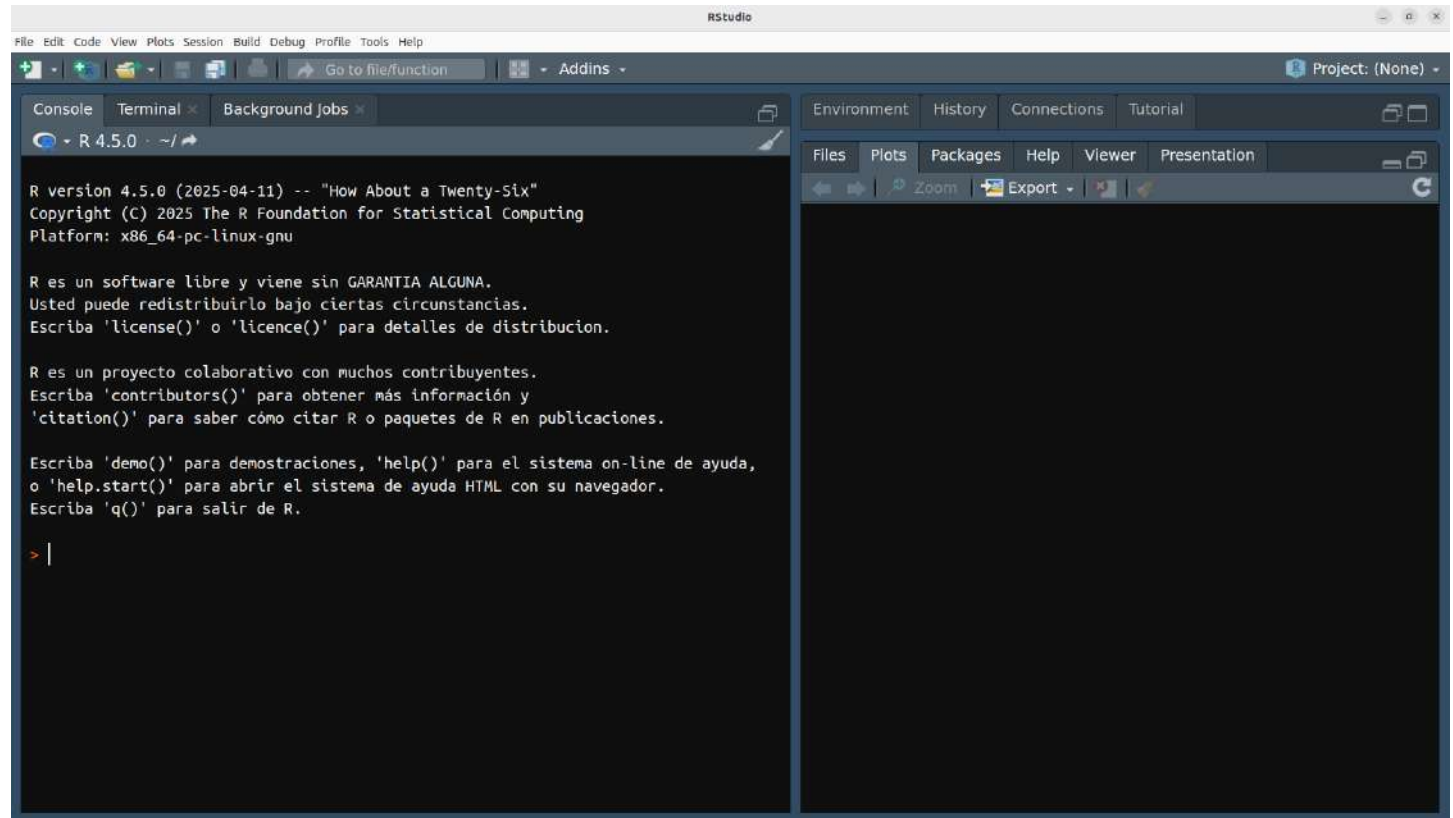
- También puede usarse un pseudo  $R^2$  que, basado en la devianza, informa sobre la ganancia de poder predictivo de tu modelo respecto al modelo nulo:

$$R^2_{\text{McFadden}} = 1 - \frac{D_{\text{modelo}}}{D_{\text{nulo}}}$$

- $D_{\text{modelo}}$ : deviance del modelo ajustado
- $D_{\text{nulo}}$ : deviance del modelo con solo intercepto

Valor de $R^2_{\text{McFadden}}$	Interpretación aproximada
0.0 – 0.2	Ajuste pobre
0.2 – 0.4	Ajuste razonable
0.4 – 0.6	Ajuste bueno
> 0.6	Ajuste excelente

## El modelo lineal: bondad de ajuste

A screenshot of the RStudio console window. The title bar says 'RStudio'. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar has icons for running code, saving, and other functions. The console shows the R version 4.5.0 (2025-04-11) and copyright information. It also displays instructions for using R, such as 'R es un software libre y viene sin GARANTIA ALGUNA.' and 'Escriba 'license()' o 'licence()' para detalles de distribucion.' The prompt '>' is visible at the bottom of the console.

```
R version 4.5.0 (2025-04-11) -- "How About a Twenty-Six"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> |
```